RESSALVA

Atendendo solicitação do autor, o texto completo desta Tese será disponibilizado somente a partir de 28/06/2026.

UNIVERSIDADE ESTADUAL PAULISTA
"JÚLIO DE MESQUITA FILHO"
Câmpus de Rio Claro

Programa de Pós-Graduação em Ciência da Computação

Lucas Pascotti Valem

Contextual Similarity Learning for Image Retrieval and Classification: Applications in Person Re-Identification

Rio Claro - SP

2024

UNIVERSIDADE ESTADUAL PAULISTA
“Júlio de Mesquita Filho”
Instituto de Geociências e Ciências Exatas
Câmpus de Rio Claro

Lucas Pascotti Valem

Contextual Similarity Learning for Image Retrieval and Classification: Applications in Person Re-Identification

Tese de Doutorado apresentada ao Instituto de Geociências e Ciências Exatas do Câmpus de Rio Claro, da Universidade Estadual Paulista “Júlio de Mesquita Filho”, como parte dos requisitos para obtenção do título de Doutor em Ciência da Computação.

Orientador: Prof. Dr. Daniel Carlos Guimarães Pedronette

Rio Claro - SP
2024

V151c

Valem, Lucas Pascotti

Contextual similarity learning for image retrieval and classification: applications in person re-identification / Lucas Pascotti Valem. -- Rio Claro, 2024

261 f. : il., tabs.

Tese (doutorado) - Universidade Estadual Paulista (UNESP), Instituto de Geociências e Ciências Exatas, Rio Claro

Orientador: Daniel Carlos Guimarães Pedronette

1. Ciência da Computação. 2. Inteligência Artificial. 3. Recuperação de Imagens por Conteúdo. 4. Classificação de Imagens. 5. Re-Identificação de Pessoas. I. Título.

Sistema de geração automática de fichas catalográficas da Unesp. Biblioteca da Universidade Estadual Paulista (UNESP), Instituto de Geociências e Ciências Exatas, Rio Claro. Dados fornecidos pelo autor(a).

Essa ficha não pode ser modificada.

UNIVERSIDADE ESTADUAL PAULISTA
“Júlio de Mesquita Filho”
Instituto de Geociências e Ciências Exatas
Câmpus de Rio Claro

Lucas Pascotti Valem

Contextual Similarity Learning for Image Retrieval and Classification: Applications in Person Re-Identification

Tese de Doutorado apresentada ao Instituto de Geociências e Ciências Exatas do Câmpus de Rio Claro, da Universidade Estadual Paulista “Júlio de Mesquita Filho”, como parte dos requisitos para obtenção do título de Doutor em Ciência da Computação.

Comissão Examinadora

- Prof. Dr. Daniel Carlos Guimarães Pedronette (Orientador)
  Instituto de Geociências e Ciências Exatas (IGCE)
  Universidade Estadual Paulista - UNESP
- Profa. Dra. Agma Juci Machado Traina
  Instituto de Ciências Matemáticas e de Computação (ICMC)
  Universidade de São Paulo - USP
- Prof. Dr. Hélio Pedrini
  Instituto de Computação (IC)
  Universidade Estadual de Campinas - UNICAMP
- Prof. Dr. João Paulo Papa
  Faculdade de Ciências (FC)
  Universidade Estadual Paulista - UNESP
- Prof. Dr. Wallace Correa de Oliveira Casaca
  Instituto de Biociências, Letras e Ciências Exatas (IBILCE)
  Universidade Estadual Paulista - UNESP

Conceito: Aprovado.
Rio Claro (SP), 28 de junho de 2024.

Acknowledgements

First and foremost, I thank God for the gift of life and health.

My parents and family, for their love and unconditional support.

My advisor, for the guidance, support, and trust.

All the professors, both national and international, who contributed to this research.

The university and all the faculty members of the graduate program.

Friends and colleagues, for the encouragement and support.

The São Paulo Research Foundation (FAPESP), Fulbright, and Petrobras for the financial support.

Resumo

O crescimento exponencial das coleções de imagens produziu um aumento significativo nas aplicações de aprendizado de máquina e recuperação de imagens em diversos cenários. Apesar dos avanços recentes, muitos métodos ainda dependem fortemente de grandes volumes de dados rotulados para treinamento, o que representa um obstáculo importante, uma vez que produzir dados rotulados é geralmente custoso. Para enfrentar esse desafio, várias técnicas foram desenvolvidas. Um aspecto crítico de tais abordagens é definir a similaridade entre imagens de maneira eficaz, o que continua sendo um desafio central em aplicações de recuperação e aprendizado de máquina, tais como classificação. A questão central está intrinsecamente relacionada à forma como a informação é representada e aos métodos usados para comparar essas representações. Uma grande limitação é que a maioria ainda depende de medidas par-a-par e ignoram outras informações significativas presentes na vizinhança que podem ser usadas para melhorar os resultados. Este trabalho foca em melhorar a eficácia da recuperação de imagens por conteúdo visual e tarefas de classificação usando similaridade contextual, indo além das métricas tradicionais par-a-par para explorar as relações entre os elementos. O aprendizado de similaridade contextual é empregado para explorar relações de vizinhança entre os elementos, usando técnicas tais como informações baseadas em ranqueamento, medidas contextuais, grafos e hipergrafos para modelar a informação contextual de forma eficaz. Esta tese propõe sete métodos novos aplicados a cenários de propósito geral e re-identificação de pessoas (Re-ID) abordando diferentes contribuições. Três tarefas principais foram consideradas: estimativa de eficácia de consultas, recuperação e classificação de imagens. Foi realizada uma ampla avaliação experimental, totalizando 17 coleções de imagens e mais de 50 descritores visuais. Os métodos propostos, quando comparados com o estado-da-arte, demonstram resultados que são comparáveis ou superiores aos das abordagens existentes na maioria dos casos.

Palavras-chave: Similaridade Contextual; Recuperação de Imagens; Classificação de Imagens; Estimativas de Eficácia; Re-identificação de Pessoas; Aprendizado de Representações.

Abstract

The exponential growth of image collections has demanded a significant increase in the use of machine learning and image retrieval applications across various scenarios. Despite the relevant advances, many methods still rely heavily on large volumes of labeled data for training, which establishes an important obstacle, once producing labeled data is generally expensive and time-consuming. To address this challenge, numerous techniques have been developed recently. A critical aspect of these approaches is effectively defining image similarity, which remains a central challenge in retrieval and machine learning applications, such as classification. The core of this issue is intrinsically linked to how information is represented and the methods used to compare these representations. A major limitation is that most of them still rely on pairwise measures, ignoring other meaningful information present in the neighborhood that can be used to further increase the results. This work focuses on improving the effectiveness of image retrieval by visual content and classification tasks using contextual similarity, moving beyond traditional pairwise measures to exploit relationships among elements. Contextual similarity learning is employed to capture underlying relationships among elements, using techniques such as rank-based models, contextual measures, graphs, and hypergraphs to model contextual information effectively. This dissertation proposes seven novel methods applied across general-purpose and person re-identification (Re-ID) scenarios addressing different contributions. Three main tasks were considered: query performance prediction, image retrieval, and image classification. A wide experimental evaluation was conducted, totaling 17 datasets and more than 50 visual image descriptors. The proposed methods, when compared with state-of-the-art and recent baselines, demonstrate results that are comparable to or surpass those of existing approaches in most cases.

Keywords: Contextual Similarity Information; Image Retrieval; Image Classification; Query Performance Prediction; Person Re-ID; Representation Learning.

List of Figures

List of Tables

List of Abbreviations and Acronyms

ACC	Color Autocorrelogram Descriptor
ACF	Aggregated Channel Features Detector
AF	Attention Features
AIN	Adaptive Instance Normalization
AIR	Articulation-Invariant Representation Descriptor
ALOI	Amsterdam Library of Object Images
ANML	Adaptive Neighborhood Metric Learning
AP	Average Precision
APPNP	Approximate Personalized Propagation of Neural Predictions
ARMA	Auto-Regressive Moving Average Filter Convolution
ARN	Adaptation and Re-Identification Network
ASC	Aspect Shape Context Descriptor
ATNET	Adaptive Transfer Network
BAS	Beam Angle Statistics Descriptor
BFS	Breadth-First Search
BFSTREE	Breadth-First Search Tree
BIC	Border/Interior Pixel Classification Descriptor
BOVW	Bag of Visual Words
BOW	Bag of Words
CAMEL	Cross-view Asymmetric Metric LEarning
CAP	Camera-aware proxies
CBIR	Content-Based Image Retrieval
CC	Connected Components
CCL	Contextual Contrastive Loss
CCOM	Color Co-Occurrence Matrix Descriptor
CEDD	Color and Edge Directivity Descriptor
CFD	Contour Features Descriptor
CG	Correlation Graph
CIFAR	Canadian Institute For Advanced Research
CLD	Color Layout Descriptor
CMC	Cumulative Matching Characteristics
CNN	Convolutional Neural Networks
COMO	Compact Composite Moment-Based Descriptor
CPRR	Cartesian Product of Ranking References
CPU	Central Processing Unit
CSGLP	Camera Style Generation and Label Propagation
CSRT	Discriminative Correlation Filter with Channel and Spatial Reliability
CUB200	Caltech-UCSD Birds Dataset
CUHK	Dataset from the Chinese University of Hong Kong
DAAM	Domain Adaptive Attention Model
DCNN	Deep Convolutional Neural Networks
DIDAL	Discriminative Identity-Feature Exploring and Differential Aware Learning

DPM	Deformable Parts Model
DPNET	Dual Path Network
DRNE	Deep Rank Noise Estimator
DUKEMTMC	Duke Multi-Tracking Multi-Camera Dataset
EANET	Enhancing Alignment Network
ECN	Exemplar Memory Convolutional Network
EHD	Edge Histogram Descriptor
ELF	Ensemble of Localized Features
EMTL	Enhanced Multi-Dataset Transfer Learning
FBRESNET	Facebook Residual Neural Network
FCTH	Fuzzy Color and Texture Histogram
FN	False Negatives
FOH	Fuzzy Opponent Histogram
FP	False Positives
GAN	Generative Adversarial Network
GAT	Graph Attention Networks
GB	Gigabytes
GBICOV	Covariance descriptor based on Bio-inspired Features
GCN	Graph Convolutional Network
GCN-APPNP	Approximate Personalized Propagation of Neural Predictions GCN
GCN-ARMA	Auto-Regressive Moving Average Filter Convolution GCN
GCN-GAT	Graph Attention Networks GCN
GCN-SGC	Simple Graph Convolution GCN
GDP	Graph Diffusion Process
GIST	Global Image Descriptor for low-dimensional features
GNN	Graph Neural Network
GNN-KNN-LDS	KNN variation of GNN-LDS
GNN-LDS	Learning Discrete Structures for Graph Neural Networks
GOG	Gaussian Of Gaussians Descriptor
GPU	Graphics Processing Unit
GRAD-NET	Graph Diffusion Network
GRID	UnderGround Re-IDentification (GRID) Dataset
GS	Graph Sampling
GSP	Graph Signal Processing
GSSL	Graph-based Semi-Supervised Learning
HACNN	Harmonious Attention Network
HCT	Hierarchical Clustering with Hard-batch Triplet Loss
HHL	Hetero and Homogeneously Learning
HLBP	Histogram of Local Binary Patterns
HQPP	Hypergraph Query Performance Prediction
HRSF	Hypergraph Rank Selection and Fusion
HSV	Color Space: Hue, Saturation, Value
IBN	Instance-Batch Normalization
ICE	Inter-Instance Contrastive Encoding
ICS	Intra-Camera Supervise
ID	Identifier
IDSC	Inner Distance Shape Context
IICS	Intra-inter Camera Similarity

IR	Information Retrieval
ISSDA	Iterative Self-Supervised Domain Adaptation
JCD	Joint Composite Descriptor
JVCT	Joint Generative and Contrastive Learning
KCF	Kernelized Correlation Filters
KISSME	Keep-it-simple-and-straightforward distance learning
KNN	K Nearest Neighbors
LAS	Local Activity Spectrum
LBP	Local Binary Patterns
LCDP	Locally constrained diffusion process
LDA	Linear Discriminant Analysis
LDFV	Local Descriptors encoded by Fisher Vector
LDS-GNN	Learning Discrete Structures for Graph Neural Networks
LGBM	Light Gradient Boosting Machine
LHRR	Log-based Hypergraph of Ranking Reference
LMNN	Large margin nearest neighbor learning
LOMO	Local Maximal Occurrence Descriptor
LS	Label Spreading
LSTM	Long short-term memory
MAM	Memory Access Method
MAP	Mean Average Precision
MAR	MultilAbel Reference Learning
MATE	Multi-Task Multi-Label
MCFS	Muti-cluster Feature Selection
MCRN	Multi-Centroid Representation Network
MGCE-HCL	Multi-Granularity Clustering Ensemble-based Hybrid Contrastive Learning
MGH	Metadata Guided Hypergraph
ML	Machine Learning
MLFN	Multi-Level Factorisation Network
MMCL	Memory-based Multi-label Classification Loss
MOSSE	Minimum Output Sum of Squared Error
MPEG	Moving Picture Experts Group
MR	Manifold Ranking
MSE	Mean Squared Error
MSMT	Multi-Scene Multi-Time Re-ID Dataset
NASNET	Neural Architecture Search Network
NDFS	Non-negative Discriminative Feature Selection
NET	Network
NMF	Non-negative Matrix Factorization
NNCLR	Nearest-Neighbor Contrastive Learning of Visual Representations
NP-HARD	Nondeterministic Polynomial-time Hard
NS	Abbreviation of NS-Score
NS-SCORE	Score for UKBench Dataset
O2CAP	Offline-Online Associated Camera-Aware Proxies
OLDFP	Object Level Deep Feature Pooling
OOD	Out-of-Domain
OPF	Optimum-Path Forest

ORL	Our Database of Faces
OSNET	Omni-Scale Feature Learning Neural Network
OSNET-AIN	OSNET with Adaptive Instance Normalization
OSNET-IBN	OSNET with Instance-Batch Normalization
PAF	Part Association Field
PAUL	Patch-Based Unsupervised Learning Framework
PCA	Principal Component Analysis
PHOG	Pyramidal Histogram of oriented gradients
PIF	Part Intensity Field
PK-SAMPLER	Random Sampling Method in Re-ID
QPP	Query Performance Prediction
RAM	Random Access Memory
RBF	Radial Basis Function Kernel
RBO	Rank-Biased Overlap
RDNN	Residual Dense Neural Network
RDP	Regularized Diffusion Process
RDPAC	Rank Diffusion Process with Assured Convergence
RE-ID	Person Re-Identification
RESNET	Residual neural network
RFE	Rank Flow Embedding
RGB	Red, Green, Blue
RL-SIM	Ranked Lists Similarity Approach
RLCC	Refining Pseudo Labels with Clustering Consensus
RQPPF	Regression for Query Performance Prediction Framework
SCC	Strongly Connected Components
SCD	Scalable Color Descriptor
SCH	Simple Color Histogram
SD	Self-diffusion for Image Segmentation and Clustering
SDC	Scale-invariant Feature Transform Dense Color
SENET	Squeeze-and-Excitation Network
SGC	Simple Graph Convolution
SGD	Stochastic Gradient Descent
SIFT	Scale-invariant Feature Transform
SORT	Simple Online and Realtime Tracking
SOTA	State-of-the-art
SP	Spatial Pyramid
SPACC	Spatial Pyramid Color Autocorrelogram Descriptor
SPCEDD	Spatial Pyramid Color and Edge Directivity Descriptor
SPEC	Spectral Regression
SPFCTH	Spatial Pyramid Fuzzy Color and Texture Histogram
SPGAN	Similarity preserving generative adversarial network
SPJCD	Spatial Pyramid Joint Composite Descriptor
SPLBP	Spatial Pyramid Local Binary Patterns
SS	Segment Saliences
SSL	Softened Similarity Learning Approach
STF	Swin-Transformer
SURF	Speeded-Up Robust Features
SVD	Singular Value Decomposition

SVM	Support Vector Machines
SVR	Support Vector Regression
SWIN-TF	Swin-Transformers
T-SNE	t-Distributed Stochastic Neighbor Embedding
TAUDL	Tracklet Association Unsupervised Deep Learning
TN	True Negatives
TP	True Positives
TPG	Tensor Product Graph
UDA	Unsupervised Data Augmentation
UDLF	Unsupervised Distance Learning Framework
UGAF-RSF	Unsupervised Genetic Algorithm Framework for Rank Selection and Fusion
UKBENCH	University of Kentucky Dataset
USRF	Unsupervised Selective Rank Fusion Method
UTAL	Unsupervised Tracklet Association Learning
VAL-PAT	Framework for Transferable Representations of Pedestrians
VGGNET	Visual Geometry Group Network
VIT	Vision Transformer
VOC	Vocabulary Tree
VRAM	Video Random Access Memory
WHOS	Weighted Histograms of Overlapping Stripes
WSEF	Weakly Supervised Experiments Framework
YOLO	You Only Look Once object detection network

List of Symbols

$A(i)$	Set of all elements in a batch, except the image of index i.
C	Number of virtual classes in the synthetic scenario.
C_n	Set of combinations where each combination is of size n.
E	Set of edges of a graph.
E_h	Set of hyperedges of a hypergraph.
G	A graph.
H	A hypergraph model or the number of feature maps (or hidden units) in the hidden layer of a GCN.
I	The set of indices for all augmented samples in a batch.
L	Size of ranked lists.
M	Confusion matrix of probabilities between classes.
M_c	Confusion matrix of probabilities between elements of the same class.
M_f	Sparse matrix used by RFE to accumulate normalized scores from different rankers.
N	The size of the collection $\mathcal{C}$, i.e., dataset size.
$NN_k(i)$	The set of k nearest neighbors of image i.
$NN_k^{\mathcal{Y}}(i)$	A subset of $NN_k(i)$ containing only images from the same class of image i.
N_b	Number of image pairs in a training batch.
$P(i)$	The set of indices of all positive samples in the batch distinct from image i.
R_i	Ranker of index i.
S	Selection set of all possible combinations of rankers.
S_p	Selection set of pairs of rankers.
T, t	Number of iterations.
V	Set of graph vertices.
V_L	Set of labeled nodes in the graph.
V_U	Unlabeled subset of the node set.
α	Constant for the normalization equation of RFE.
β	Weight or relevance of correlation in the selection measure.
$\boldsymbol{z}_i$	The embedding of the data sample i generated by the metric learning model.
$\circ$	Hadamard (element-wise) product.
δ	A function that computes the distance between two feature vectors.
ϵ	A function that extracts a feature vector from an image.
η_f	Fused affinity measure used for rank aggregation.
$\eta_r(i,x)$	Function that assigns a weight to image x according to its position in τ_i.
γ	An effectiveness estimation measure.
γ_A	Authority effectiveness estimation measure.
γ_R	Reciprocal Density effectiveness estimation measure.
$\hat{\mathbf{A}}$	Normalized adjacency matrix of a graph.

λ	A correlation measure (e.g., RBO).
$\mathbb{R}$	The set of real numbers.
$\mathbf{A}$	Affinity matrix (RFE) or adjacency matrix (GCNs).
$\mathbf{C}$	Similarity measure matrix based on Cartesian product.
$\mathbf{D}$	Distance matrix.
$\mathbf{H_G}$	HRSF Hypergraph model.
$\mathbf{H}$	Incidence matrix for HRSF or matrix encoding the similarity information of h-embeddings for RFE.
$\mathbf{I}$	Identity matrix.
$\mathbf{S}$	Similarity matrix.
$\mathbf{W}$	Affinity matrix (HRSF) or weight matrix in the definition of GCNs.
$\mathbf{X}$	Feature vectors provided as input to the GCN.
$\mathbf{Z}$	Matrix of embeddings learned by the GCN model.
$\mathbf{b}$	Reciprocal neighborhood binary vector used in the computation of RQPPF meta-features.
$\mathbf{b_i}$	Reciprocal neighborhood binary vector for image i.
$\mathbf{c}_i$	Connected component of index i.
$\mathbf{c}_q$	CC-embedding of a connected component q computed by RFE.
$\mathbf{e}_i$	Representation vector (embedding) of the element of index i from the dataset.
$\mathbf{f}$	Contextual Rank-based Feature (Meta-Feature) vector.
$\mathbf{f}_s$	Set of synthetic features used for training the regression model.
$\mathbf{f}_t$	Set of test features used for testing the regression model.
$\mathbf{h}_i$	Row i of matrix $\mathbf{H}$, named h-embedding.
$\mathbf{p}$	Reciprocal rank position vector used in the computation of RQPPF meta-features.
$\mathbf{p_i}$	Reciprocal rank position vector for image i.
$\mathbf{q}$	Effectiveness estimation vector used in the computation of RQPPF meta-features.
$\mathbf{q_i}$	Effectiveness estimation vector for image i.
$\mathbf{x}_i$	Feature vector representing the image o_i, a row of matrix $\mathbf{X}$.
$\mathbf{z}_i$	Row i of matrix $\mathbf{Z}$, embedding representation for the node v_i.
$\mathcal{C}$	Image dataset.
$\mathcal{C}_L$	Set containing the L most similar images to image o_q in the collection $\mathcal{C}$.
$\mathcal{E}_c$	Set of candidate edges defined by RFE.
$\mathcal{L}_{\text{ccl}}$	Proposed contextual contrastive loss.
$\mathcal{L}^{\text{sup}}$	Supervised contrastive loss.
$\mathcal{N}$	Neighborhood set.
$\mathcal{N}(o_q, k)$	Neighborhood set containing the k most similar elements to o_q.
$\mathcal{N}_r(o_q, k)$	Reciprocal neighborhood set for image o_q.
$\mathcal{N}_r$	Reciprocal neighborhood set.
$\mathcal{R}$	Set of rankers.
$\mathcal{S}$	Set of connected components.
$\mathcal{T}$	Set of ranked lists for all the images in the dataset.
$\mathcal{T}_i$	Set of ranked lists produced by ranker R_i.
$\mathcal{T}_j$	The set of ranked lists produced by the ranker R_j.

$\mathcal{X}$	A subset of $\mathcal{C}$.
$\mathcal{Y}$	A set of labels (classes).
$\mathfrak{R}$	Set of rankers provided as input to the method.
$\mathfrak{X}^*$	Selected combination among all sizes.
$\mathfrak{X}_n^*$	Selected combination composed by n rankers.
$\mathfrak{X}_n$	Candidate combination composed by n rankers.
μ	Constant used in RBO correlation measure.
ϕ	Regression model for query performance prediction.
ψ	The Contrastive loss temperature parameter.
ρ	A similarity measure.
σ	Normalization function used in RFE.
τ_n^R	Ordered list of combinations of size n. Also referred to as the selection list.
$\tau_n^R(\mathfrak{X}_n^i)$	Position of the combination $\mathfrak{X}_n^i$ in the selection list τ_n^R.
τ_q	Ranked list of image q.
$\tau_q(i)$	The position of image o_i in the ranked list τ_q.
$\tau_{i,q}$	Ranked list of image of index q calculated by ranker i.
τ_i	Ranked list of image of index i.
$\tau_{q,f}(i)$	Position of image o_i in the ranked list of o_q according to feature f.
$\tau_q(i)$	Position of the image i in the ranked list of query image q.
$\boldsymbol{x}_\ell$	The ℓ-th image in the batch.
$\tilde{\boldsymbol{x}}_\ell$	The ℓ-th augmented image in the batch.
$\boldsymbol{y}_\ell$	The label corresponding to the ℓ-th image.
$\tilde{\boldsymbol{y}}_\ell$	The label corresponding to the ℓ-th augmented image.
$\tilde{\mathbf{A}}$	Adjusted adjacency matrix, $\mathbf{A} + \mathbf{I}$.
$\tilde{\mathbf{D}}$	Degree matrix of $\tilde{\mathbf{A}}$.
$\times$	Multiplication operator.
ξ	The current epoch number.
ξ_{total}	The total number of epochs.
a_{ij}	Entry in the adjacency matrix indicating the presence (1) or absence (0) of an edge between vertices o_i and o_j.
c	Number of classes (or categories).
$c(i,j)$	Element of matrix $\mathbf{C}$.
cp	Pairwise similarity relationship based on Cartesian product.
d	Number of vector dimensions.
d_e	The dimensionality of the RFE embedding space in which each object is represented.
e_i	A hyperedge of index i.
f_g	Function that, given a hypergraph and an incidence matrix, calculates a graph (RFE).
f_h	Function that, given ranked lists, calculates a hypergraph and an incidence matrix by re-ranking through hypergraph embeddings (RFE).
f_m	Manifold learning function that processes a set of ranked lists $\mathcal{T}$.
$f_p(o_q, i)$	Function returning the i-th neighbor of image q.
f_r	Function representing unsupervised similarity learning.
f_s	Function for ranker selection.
f_{gcn}	Function representing the graph convolutional network model.

$h(e_i, v_j)$	Reliance of vertex v_j to belong to a hyperedge e_i.
$h_p(e_q)$	Weight of hyperedge e_q.
h_{ij}	An element of $\mathbf{H}$ representing the similarity of object o_j in the context of the hyperedge e_i.
k	Size of the neighborhood set.
k_d	A variable representing a specific depth for computing a correlation measure.
k_{start}	The initial value of k for the first epoch.
k_v	Size of virtual classes for synthetic data.
m	Number of features, i.e., size of the set $\mathfrak{R}$.
n	Size of a combination.
n_k	Number of candidate edges for RFE graph.
o_i	Indicates any object (element) belonging to the dataset, whose index is i.
obj_i	Object of index i, often abbreviated as o_i.
p	Pairwise relationship function defined by RFE.
p_x	Pixel of position x in a grayscale image.
s_c	RFE Similarity measure attributed to pairs based on the similarity between h-embeddings and confidence of the hyperedge.
t_c	Threshold for edge computation in the connected components stage of the RFE.
th_{end}	Final threshold of the Correlation Graph.
th_{inc}	Correlation Graph threshold increment.
th_{start}	Initial threshold of the Correlation Graph.
v_i	A node in the node set V representing an image o_i.
v_l	A labeled node.
w	Selection measure for combinations of rankers.
$w(e_i)$	A positive weight assigned to a hyperedge e_i.
$w_p(i, x)$	A weight function that assigns relevance to a vertex o_x based on its position in a ranked list.
w_p	Selection measure for pairs of rankers proposed by HRSF.
$\mathcal{T}_h^{(T)}$	Set of ranked lists after T iterations of RFE.
y_i	Label (class) of object o_i, i.e., i-th row of $\mathcal{Y}$.

Contents

1 Introduction

Effectively defining the similarity between images is a central challenge in retrieval and machine learning applications. This issue is deeply connected to: *(i)*: how information is represented; and *(ii)*: the measures used to compare these representations [321, 294, 371, 250]. This work presents contributions in both directions, proposing seven novel approaches.

This dissertation discusses and presents contributions aimed at improving the effectiveness of image retrieval by visual content and classification tasks using contextual similarity learning. This introductory chapter outlines an overview of this work and is organized as follows: Section 1.1 discusses the motivations of the conducted research. Section 1.2 presents the challenges and research questions addressed. Section 1.3 states the main hypothesis validated in this dissertation. Section 1.4 discusses the objectives and contributions of the study. Section 1.5 describes the overall structure of this document, including a summary of each chapter's content and an illustration of how concepts and terms relate to the contributions.

1.1 Motivation

In recent years, there has been an exponential increase in the volume of image data, primarily due to advancements in technologies for generating, storing, and sharing visual information [320, 80]. Additionally, there are numerous applications (e.g., surveillance cameras [390, 137, 426], medical imaging [341, 6, 1], remote sensing systems [160], social media [140]) that generate vast amounts of visual data.

In this scenario, image retrieval and machine learning tasks such as image classification are increasingly being utilized in many applications [255]. Remarkable progress has been made in these methods, particularly due to the consistent evolution of deep learning [109, 31]. However, the majority of them are supervised and depend on large volumes of labeled data for training. In contrast, the production of labeled data is challenging since it is often expensive and time-consuming to obtain [91]. It may also require a specialist for labeling, especially according to the specificity of the domain. Aiming at filling this gap, many unsupervised, semi-supervised, and even self-supervised approaches have been proposed to deal with such a challenge [107]. In most of these methods, effectively modeling data is crucial for exploiting the information available in the unlabeled data.

For most approaches, the essence of learning hinges on the ability to model data accurately, which involves different concepts, in particular, representation approaches and distance or similarity measures [321]. This is especially important for Content-Based Image Retrieval (CBIR) systems, which retrieve images based on visual content rather than metadata [316]. These systems usually employ feature extraction and representation methods, which have evolved considerably [294]. Such methods have transitioned from traditional hand-crafted features [254] to more advanced deep learning approaches [294, 438], including Vision Transformers [77, 202]. However, most comparison tasks still rely on pairwise measures [294, 80], which do not exploit contextual information [245]. In general, the term context can be broadly understood as all the relevant information pertinent to an application and its users. This work considers the idea of *contextual similarity* that consists of exploiting the relationships beyond pairwise analysis, involving other elements, such as the neighborhood or more related additional information [246, 245]. The term *contextual similarity learning* is used for the learning process that employs contextual similarity for more effectively capturing the underlying relationships among elements.

An essential aspect is that contextual similarity information can be modeled in many different forms, using different representations and structures [235, 416], among them: *(i)* graphs [86, 366]: they can be used for exploiting the relationships between neighbors, which is a key aspect for understanding the local context and influence among interconnected entities; *(ii)* ranked lists [246, 245]: in image retrieval, each ranked list contains the most similar elements for a given query. The similarities between elements can be redefined according to the analysis of the neighborhoods available in these lists. The position of each element in each list also contains valuable information; *(iii)* clustering [404, 373]: identify and group data points that are similar according to predefined criteria. These groups allow the discovery of inherent patterns or relationships that may not be apparent upon initial observation.

Besides these approaches, there is still a wide range of methodologies that can be proposed to exploit contextual information in numerous scenarios. Similarity learning applied to retrieval is generally explored by re-ranking tasks. Despite the crescent popularity of these methods, more robust structures have not yet been extensively employed in most cases. Structures that represent higher-order similarities, such as relationships among neighbors of neighbors, can be particularly advantageous. Hypergraphs, for example, allow edges to connect multiple vertices, offering a sophisticated technique for capturing these relationships [403, 251]. Additionally, most unsupervised re-ranking approaches [18, 282, 108] provide a new ranked list representation as output, but new features are not produced in return, which could be used to encode contextual information for classifiers, for example.

Another application that could deeply take advantage of the enrichment of contextual information is feature selection and fusion [260, 389, 424, 329, 327]. Among different strategies, the selection of features can be done through effectiveness estimation and correlation measures [329]. They consider the idea that fusion benefits from elements with high effectiveness and that are also complementary. The creation and usage of contextual structures for estimating the effectiveness and measuring correlations is still an area that requires further research.

This work exploits contextual similarity learning for general-purpose image retrieval and person re-identification, usually abbreviated as person Re-ID. Person Re-ID is a type of surveillance application that has been gaining a lot of attention and nowadays is of fundamental importance in many camera surveillance systems. The task consists of identifying individuals across multiple cameras that have no overlapping views [137]. A Re-ID system broadly consists of three main steps [426]: person detection, feature extraction, and person retrieval or matching. This work focuses on the final step, which can be viewed as a specific image retrieval application [137].

Person Re-ID is a complex task that presents numerous difficulties [390, 137, 426], including *(i)* varying angles of view between cameras, *(ii)* low-resolution images, *(iii)* changes in lighting conditions, *(iv)* occlusions blocking part of the view, *(v)* the difficulty of manually labeling images for use in training algorithms, *(vi)* unbalanced classes or classes with very few elements, *(vii)* the complexity of modeling data, and *(viii)* the extensive volume of data that needs to be processed.

Amid these challenges in Re-ID, many approaches introduced more robust deep learning models [390], such as Vision Transformers [111, 163, 221], metric learning [185, 156, 185, 396], and Siamese networks [340, 339]. Other strategies include dataset expansion with augmentations [132, 291] or artificial data considering appearance attributes, body parts, temporal information, and different types of multimodal information. Additionally, metric learning is often employed for Re-ID due to its capacity to be effective when dealing with unseen data [185] since it focuses on learning distances or similarities rather than specific features of the training data. This approach allows the model to generalize better to new examples that were not present in the training set.

Apart from all these advancements, post-processing methods that exploit contextual information have gained significant attention due to their ability to improve results provided by latent features of different deep learning models. Various unsupervised post-processing strategies are based on the idea of exploiting the information of reciprocal neighborhoods and measuring the co-occurrence of elements in ranked lists [429, 174, 225, 165, 211, 96, 388, 95, 108], demonstrating substantial improvements. Although these approaches are becoming increasingly common in Re-ID, methods for selection and fusion remain relatively scarce. This scenario highlights the importance of investigating methods capable of effectively exploiting contextual information.

In addition to retrieval, it is also imperative to address scenarios with limited labeled data for classification [135]. Graph convolutional networks (GCNs) offer a promising solution for semi-supervised classification by learning from both labeled and unlabeled data considering graph structures [412]. Moreover, GCNs can learn node and graph embeddings that capture complex dependencies and structural relationships [141]. However, GCNs are not widely used for image classification since graphs are typically not available in image domains [274, 343, 307]. Therefore, effectively modeling these graphs, which can be utilized to exploit contextual information, is a crucial topic for research.

Another approach that has recently demonstrated continuous advances for improving classification results is the use of contrastive learning [143, 47]. Unlike the commonly used cross-entropy loss, which aims to minimize the difference between the predicted and true class probabilities, contrastive loss focuses on learning similarities and dissimilarities between data points rather than merely categorizing them [47]. Despite this, most contrastive losses consider only pairwise measures [143, 47, 49, 47], with only a few incorporating some type of neighborhood information [441, 82, 183]. Moreover, these approaches often require huge volumes of data for training (i.e., labeled or unlabeled) [47, 49], even in self-supervised scenarios, which is a challenge in circumstances where data is scarce.

In light of the presented discussion and all challenges, the focus of this dissertation is to exploit the use of contextual similarity information with the objective of improving the effectiveness of image retrieval and classification, particularly in cases where labeled data is limited or non-existent. This dissertation primarily concentrates on unsupervised learning, while also proposing semi-supervised and supervised approaches.

1.2 Research Challenges

Contextual similarity information can be applied in a variety of fields. However, appropriately representing and exploiting contextual information in each scenario poses significant difficulties. There are many research challenges related to various applications that can be used to improve the effectiveness of image retrieval and classification tasks. In the following, several topics are discussed and corresponding research questions are presented for each:

- **Selection and fusion in person Re-ID:** The selection and fusion involves choosing the most relevant features from the data and combining them to enhance the retrieval effectiveness [260]. There are various feature extractors and possible combinations between them. Selecting the right features is crucial because manually evaluating all combinations becomes impractical as the number of features increases linearly and the number of combinations increases exponentially. The concept is based on the idea that fusion is most effective when it involves elements that are both highly efficient and complementary. For person Re-ID, the complexity of accurately matching

individuals across different camera views becomes significantly more challenging in unsupervised applications due to the absence of labeled data [137]. Effectively modeling and exploiting patterns in the data is crucial in this scenario.
Research question:

- *How can contextual similarity information be used for selection and fusion in unsupervised person Re-ID?*

- **Query performance prediction:** Also known as effectiveness estimation, query performance prediction (QPP) encompasses techniques for assessing the quality of ranked lists in scenarios where no labels are provided. In this context, the ability to assess the effectiveness of the retrieval process provides a significant advantage for different tasks, including enabling the selection of more effective ranked lists. However, QPP is very challenging, especially in unsupervised tasks. One of the main difficulties is elaborating an approach that effectively generalizes across diverse scenarios [262]. Bridging this gap represents a major challenge that can be mitigated by incorporating contextual similarity information.
Research question:

  - *How can data be modeled using contextual similarity information for query performance prediction?*

- **Synthetic data:** Recently, self-supervised approaches have been proposed to address scenarios where labeled data is scarce. Among the different means of self-supervision, one of them is by using synthetic data. There are many advantages and benefits of using synthetic data, primarily due to its flexibility and control in generating large volumes of annotated data. In domains where safety and privacy are relevant, using real data can raise privacy concerns and legal issues. Synthetic data does not carry these risks. However, creating representative synthetic data presents many difficulties. One of the primary challenges is to accurately reflect the complexity and variability of real-world data [72]. The generated synthetic data is expected to encompass a wide range of scenarios, including rare events and edge cases, to ensure comprehensive learning.
Research questions:

  - *How can contextual similarity information be used to generate synthetic data?*
  - *How can contextual similarity learning be employed on synthetically generated data?*

- **Unsupervised similarity learning methods:** Despite the potential of unsupervised similarity learning methods to improve retrieval results, effectively representing and encoding the maximum amount of contextual information remains

a challenge. This difficulty is amplified because these methods operate without labels and cannot utilize relevance feedback [312] as supervised algorithms do. These methods usually exploit the relationships among images through ranked lists and similarity among elements [95, 108, 250]. The primary challenge lies in modeling and leveraging this similarity information, which can be approached through various strategies such as graph structures [381], contextual measures [128], and more [382, 384]. Utilizing more complex structures to represent second-order similarity (i.e., relationships such as neighbors of neighbors) can be particularly relevant, for example.
Research question:

  - *How can more complex structures, which encode contextual information more effectively, be applied to unsupervised similarity learning?*

- **Representation learning and embeddings:** Feature learning is of fundamental importance in many retrieval and classification applications [321]. However, the capacity to encode information of an image in an embedding is very challenging. When converting an image into an embedding, some information is inevitably lost. This loss must be minimized to ensure that the most critical features of the image are retained. There is also the semantic gap [24, 115] between the raw pixel data of an image and the human interpretation of the image's content. Unsupervised similarity learning approaches usually post-process ranked lists to enhance image retrieval results but do not provide any form of embeddings that can be used for other tasks, such as classification.
Research question:

  - *How can contextual information from similarity learning approaches be encoded to generate embeddings that are useful for tasks beyond retrieval, such as classification?*

- **Contextual similarity and Graph Convolutional Networks (GCNs):** The GCNs effectively capture relationships and interactions within complex networks, enhancing results in tasks involving structured data. However, graphs are not inherently available for most image datasets, and GCNs heavily rely on these structures to deliver significant results [141, 307]. The main challenge involves accurately modeling the graph for effective use by the GCN.
Research question:

  - *How can contextual similarity information be incorporated into the input graph utilized by Graph Convolutional Networks (GCNs) and improve their classification results?*

- **Correlation measures and manifold learning:** Manifold learning is a technique for uncovering simpler, underlying structures in complex high-dimensional data [133]. Correlation measures quantify the similarity between data points, which is very useful to model relationships in the data. However, this is challenging since data can be complex and heterogeneous, involving multiple variables with nonlinear relationships that are difficult to capture [16]. Also, outliers may present significant challenges in data analysis.
  Research questions:
  - *Can rank-based information be utilized to measure the correlation between images more effectively?*
  - *Can a correlation measure be proposed and applied to enhance image retrieval with manifold learning?*
- **Contrastive learning**: It has been extensively used in self-supervised and supervised learning due to its effectiveness in learning representations that distinguish between similar and dissimilar images. It offers an alternative to cross-entropy by yielding more semantically meaningful image embeddings. However, most contrastive losses rely on pairwise measures to assess the similarity between elements [143, 47], ignoring more general neighborhood information that can be leveraged to enhance model robustness and generalization [441].
  Research question:
  - *How can contextual similarity information be incorporated into metric learning, including its direct integration into losses such as contrastive loss?*

The contributions presented and discussed in this work address those important research challenges.

1.3 Dissertation Statement

Driven by the challenges identified in the literature, primarily the difficulty of obtaining a large amount of labeled data and the increasing need for methods that exploit contextual information, we explore the application of contextual similarity learning in different scenarios. The main hypothesis of the work is briefly stated as follows:

> *Contextual similarity learning can improve the effectiveness of image retrieval and classification tasks across general-purpose and person re-identification (Re-ID) applications. This concept is applicable to unsupervised, semi-supervised, and supervised approaches, particularly in contexts where labeled data is limited.*

The hypothesis is validated by the proposed approaches and a comprehensive experimental evaluation presented in this dissertation.

11 Conclusions

This chapter concludes this dissertation by discussing the contributions and other relevant aspects. Section 11.1 reviews the main results obtained for each task: query performance prediction, image retrieval, and image classification. Additionally, it provides a comparative analysis of the proposed methods alongside other approaches from the literature. Section 11.2 discusses how the contributions address the research questions of this study. Section 11.3 lists the publications and submissions obtained, along with the international Fulbright fellowship. Section 11.4 mentions the available codes for the proposed approaches. Finally, Section 11.5 presents potential extensions and future work, describing their connections to the contributions achieved in this research.

11.1 Discussion of Results

Given the notable outcomes achieved by contextual similarity learning across all scenarios considered, this section discusses the results for each task. For query performance prediction, the results of RQPPF and DRNE are jointly compared and discussed in Section 11.1.1. Section 11.1.2 compares the approaches evaluated for image retrieval in person Re-ID and general-purpose datasets, including comparisons with the state-of-the-art. Section 11.1.3 overviews Manifold-GCN and RFE semi-supervised classification results obtained and a comparison with the state-of-the-art. Moreover, the gains achieved by CCL are briefly discussed.

11.1.1 Query Performance Prediction

A great variety of experiments was conducted to evaluate DRNE and RQPPF, showing their capacity to effectively perform QPP in various datasets. Table 11.1 presents a summary of the relative gains for both approaches in comparison to Authority [243] and Reciprocal Density [248], which are used as baselines. Notice that RQPPF provided gains in all the evaluated scenarios, while DRNE showed inferior performance in some cases, especially when compared to Reciprocal in the MPEG-7 dataset. However, for the most part, DRNE provided higher and more consistent gains when compared to Reciprocal. Specifically for the AIR descriptor, DRNE revealed superior results in all cases.

In general, the results showed that the proposed methods are better than the baselines in most cases. Additionally, the choice of the best method depends not only on the dataset but also on the descriptor. RQPPF is more flexible and also uses Authority and Reciprocal as part of its formulation, while DRNE does not. However, DRNE seems more robust to outlier descriptors, while RQPPF does not. Among potential extensions, combining DRNE and RQPPF is one of the possibilities for future work.

Table 11.1 – Relative gains of DRNE and RQPPF when compared to Authority (Auth.) and Reciprocal Density (Rec.). Average gains are reported for each dataset.

Descriptor	Original MAP	Compared to Auth. [243] DRNE	Compared to Auth. [243] RQPPF	Compared to Rec. [248] DRNE	Compared to Rec. [248] RQPPF
MPEG-7					
AIR [103]	89.39%	+14.81%	+3.50%	+16.84%	+12.99%
ASC [191]	85.28%	-2.50%	+3.76%	-8.29%	+1.84%
IDSC [190]	81.70%	-3.93%	+3.69%	-7.58%	+1.88%
CFD [244]	80.71%	+3.47%	+3.99%	-1.24%	+2.71%
BAS [13]	71.52%	+0.85%	+3.69%	-5.13%	+1.09%
SS [317]	37.67%	+7.08%	+6.01%	+3.32%	+3.52%
Average Gain		+3.30%	+4.11%	-0.35%	+4.01%
Brodatz					
LAS [308]	75.15%	+7.50%	+9.01%	+9.59%	+11.15%
CCOM [148]	57.57%	+3.30%	+7.33%	+8.60%	+11.18%
LBP [231]	48.40%	+0.75%	+5.10%	+18.42%	+15.36%
Average Gain		+3.85%	+7.15%	+12.20%	+12.56%
Market					
OSNET [436]	43.30%	-2.63%	+1.19%	+5.40%	+5.45%
ResNet [110]	22.82%	-0.89%	+0.13%	+7.95%	+5.29%
Average Gain		-1.76%	+0.66%	+6.68%	+5.37%
Duke					
OSNET [436]	52.69%	+0.71%	+2.35%	+1.00%	+3.37%
ResNet [110]	32.00%	-2.14%	+0.52%	-0.12%	+2.46%
Average Gain		-0.72%	+1.44%	+0.44%	+2.92%

11.1.2 Image Retrieval

Four of the seven proposed methods were evaluated in image retrieval: HRSF, JaccardMax, RFE, and Manifold-GCN. The results obtained are reviewed for both person Re-ID and general-purpose datasets, including comparisons against each other and with the state-of-the-art. A brief discussion about the gains is also presented.

- **Person Re-ID**

Considering the wide variety of descriptors employed and to provide a fair comparison, Table 11.2 presents the best results obtained for each method using only the OSNET model and its variants (i.e., OSNET, OSNET-IBN, and OSNET-AIN) on Market, DukeMTMC, and CUHK03 datasets.

For the Market and CUHK03 datasets, HRSF leads with the best results for both R1 and MAP. HRSF is the only method that performs selection, which is an advantage over the others since it can select the best combination of descriptors among the OSNET variants. For the DukeMTMC dataset, RFE and JaccardMax compete for the best results. The worst results in this table are the ones obtained by the Manifold-GCN. Besides Manifold-GCN being semi-supervised, while all the other approaches are unsupervised, this result highlights the importance of future research for this method. Since it was mainly proposed for classification, the results for retrieval are significantly behind others, probably due to the features not being properly distributed in the latent space, which requires further investigation in future work.

Table 11.3 presents the methods ranked according to their results for each measure and dataset. The average rank reveals that, while HRSF shows the best results in most cases, JaccardMax and RFE follow closely, with average ranks of 2.0 and 2.2, respectively. As previously discussed, Manifold-GCN is behind with an average rank of 3.7.

Table 11.2 – Comparison between the proposed approaches on person Re-ID considering MAP (%) and R-01 (%). The best results obtained with the OSNET descriptor and its variants are reported.

Method	Year	Datasets					
		Market1501		DukeMTMC		CUHK03	
		R1	MAP	R1	MAP	R1	MAP
HRSF ($\mathfrak{X}^*$, best result) [331]	2022	**75.71**	**62.94**	77.24	68.88	**39.04**	**39.69**
Correlation Graph + Jaccard Max [324]	2022	73.25	59.84	76.21	**69.27**	—	—
RFE [334]	2023	72.42	59.51	**77.69**	69.21	36.89	39.24
Manifold-GCN [333]	2023	70.30	57.48	74.22	65.83	35.19	35.99

Table 11.3 – Proposed approaches on person Re-ID ranked according to their effectiveness (R1 and MAP). The best results obtained with the OSNET descriptor and its variants were considered.

Method	Year	Datasets						Average Rank
		Market1501		DukeMTMC		CUHK03		
		R1	MAP	R1	MAP	R1	MAP	
HRSF ($\mathfrak{X}^*$, best result) [331]	2022	**1**	**1**	2	3	**1**	**1**	1.5
Correlation Graph + Jaccard Max [324]	2022	2	2	3	**1**	—	—	2.0
RFE [334]	2023	3	3	**1**	2	2	2	2.2
Manifold-GCN [333]	2023	4	4	4	4	3	3	3.7

To compare the proposed methods with the state-of-the-art in person Re-ID, which is presented in Table 11.4, the best results for each approach were considered. For HRSF, JaccardMax, and RFE the best results used OSNET and its variants. Unlike the others, the JaccardMax evaluation employed the TransReID descriptor, which provided better results for this method. This table highlights in bold the highest value for each column. The best among our methods is also highlighted. All the baseline results are the ones reported in the literature, following the same protocol as ours.

In general, it can be observed that the proposed approaches provide better results for MAP than R1 when compared to other methods. This evinces that they can significantly improve the top positions of ranked lists, but not necessarily achieve the best results when considering only the first position. In this case, Market1501 was revealed as the most challenging dataset, where the proposed methods are better or comparable to the ones up to 2020. After that, the baselines show a considerable improvement. In contrast, for the DukeMTMC dataset, the MAP of 73.96% obtained by the proposed JaccardMax (2022) is the best result achieved, surpassed only by VAL-PAT, which is a very recent approach from 2023. For the CUHK03 dataset, many of the methods have no results reported in

Table 11.4 – Proposed approaches compared to state-of-the-art on person Re-ID considering MAP (%) and R-01 (%).

Method	**Year**	**Datasets**					
		Market1501		**DukeMTMC**		**CUHK03**	
		R1	**MAP**	**R1**	**MAP**	**R1**	**MAP**
Unsupervised Methods							
ARN [181]	2018	70.3	39.4	60.2	33.4	—	—
EANet [118]	2018	66.4	40.6	45.0	26.4	51.4	31.7
TAUDL [170]	2018	63.7	41.2	61.7	43.5	44.7	31.2
ECN [431]	2019	75.1	43.0	63.3	40.4	—	—
MAR [397]	2019	67.7	40.0	87.1	48.0	—	—
UTAL [171]	2019	69.2	46.2	62.3	44.6	56.3	42.3
SSL [189]	2020	71.7	37.8	52.5	28.6	—	—
HCT [402]	2020	80.0	56.4	69.6	50.7	—	—
CAP [353]	2021	91.4	79.2	81.1	67.3	—	—
IICS [376]	2021	89.5	72.9	80.0	64.4	—	—
RLCC [415]	2021	90.8	77.7	83.2	69.2	—	—
ICE [43]	2021	93.8	82.3	83.3	69.9	—	—
MGH [368]	2021	93.2	81.7	83.7	70.2	—	—
MGCE-HCL [297]	2022	92.1	79.6	82.5	67.5	—	—
MCRN [367]	2022	92.5	80.8	83.5	69.9	—	—
O2CAP [354]	2022	92.5	82.7	83.9	71.2	—	—
DIDAL [201]	2023	94.2	84.8	—	—	—	—
VAL-PAT [23]	2023	—	—	86.1	74.9	—	—
Domain Adaptative Methods							
HHL (D,M) [430]	2018	62.2	31.4	46.9	27.2	—	—
HHL (C03) [430]	2018	56.8	29.8	42.7	23.4	—	—
ATNet (D,M) [197]	2019	55.7	25.6	45.1	24.9	—	—
CSGLP (D,M) [273]	2019	63.7	33.9	56.1	36.0	—	—
ISSDA (D,M) [306]	2019	81.3	63.1	72.8	54.1	—	—
ECN++ (D,M) [432]	2020	84.1	63.8	74.0	54.4	—	—
MMCL (D,M) [348]	2020	84.4	60.4	72.4	51.4	—	—
JVCT+ (D,M) [44]	2021	90.5	75.4	81.9	67.6	—	—
MCRN (D,M) [367]	2022	93.8	83.8	84.5	71.5	—	—
Cross-Domain Methods (single-source)							
EANet (C03) [118]	2018	59.4	33.3	39.3	22.0	—	—
EANet (D,M) [118]	2018	61.7	32.9	51.4	31.7	—	—
SPGAN (D,M) [71]	2018	43.1	17.0	33.1	16.7	—	—
DAAM (D,M) [121]	2019	42.3	17.5	29.3	14.5	—	—
AF3 (D,M) [195]	2019	67.2	36.3	56.8	37.4	—	—
AF3 (MT) [195]	2019	68.0	37.7	66.3	46.2	—	—
PAUL (MT) [380]	2019	68.5	40.1	72.0	53.2	—	—
Cross-Domain Methods (multi-source)							
CAMEL [396]	2017	54.5	26.3	—	—	31.9	—
EMTL [370]	2018	52.8	25.1	39.7	22.3	—	—
Baseline by [153]	2019	80.5	56.8	67.4	46.9	29.4	27.4
Proposed Methods (contributions)							
HRSF ($\mathfrak{X}^*$, best result) [331]	2022	**75.71**	62.94	77.24	68.88	**39.04**	**39.69**
Correlation Graph + Jaccard Max [324]	2022	75.42	**63.53**	**78.59**	**73.96**	—	—
RFE [334]	2023	72.42	59.51	77.69	69.21	36.89	39.24
Manifold-GCN [333]	2023	70.30	57.48	74.22	65.83	35.19	35.99

the literature, since this dataset is not as commonly evaluated as the others. However, all methods provided a better MAP than the baselines, being only behind UTAL.

The presented comparisons raise two topics for discussion: *(i)* Why the obtained results are significantly better in DukeMTMC? Why does Market1501 appear to be considerably more difficult? *(ii)* RFE and CG [249] + JaccardMax exhibit close results when using the same descriptors for Re-ID. Is this also true in other scenarios?

The first topic is challenging to answer, especially because Market1501 and DukeMTMC datasets have very similar characteristics (e.g., dataset size, number of individuals, images per person, size of the train and evaluation sets, and number of cameras). However, one particular difference might explain it. The Market1501 dataset was annotated using an automated detector, the Deformable Part Model (DPM), which is known to be prone to noise and potential misalignment. Conversely, the DukeMTMC was manually annotated by humans providing cleaner data with well-aligned bounding boxes. Further investigation to address this aspect can be conducted as future work.

Regarding the close results of RFE and CG [249] + JaccardMax for Re-ID, these methods are compared on general-purpose datasets to evaluate if they exhibit similar behavior.

- **General-Purpose Datasets**

Tables 11.5 and 11.6 compare RFE and CG [249] + JaccardMax with the state-of-the-art in image retrieval tasks for the datasets Holidays and UKBench, respectively. In both datasets, RFE outperformed all the baselines. For Holidays, CG [249] + JaccardMax is behind RFE with 91.12% but still surpasses most of the other methods. In contrast, for UKbench, both achieved the same result of 3.97, which is very close to the maximum score (i.e., 4).

Table 11.5 – State-of-the-art (SOTA) comparison on Holidays dataset (MAP).

MAP for state-of-the-art methods						
Jégou *et al.* [127]	Tolias *et al.* [315]	Paulin *et al.* [238]	Qin *et al.* [268]	Zheng *et al.* [425]	Sun *et al.* [299]	Zheng *et al.* [423]
75.07%	82.20%	82.90%	84.40%	85.20%	85.50%	85.80%
Pedronette *et al.* [241]	Arandjelovic *et al.* [12]	Li *et al.* [178]	Razavian *et al.* [271]	Pedronette *et al.* [253]	Gordo *et al.* [104]	Valem *et al.* [329]
86.16%	87.50%	89.20%	89.60%	90.02%	90.30%	90.51%
Valem *et al.* [328]	Liu *et al.* [203]	Pedronette *et al.* [251]	Pedronette *et al.* [252]	Yu *et al.* [398]	Berman *et al.* [26]	
90.51%	90.89%	90.94%	91.25%	91.40%	91.80%	

Proposed Approaches	
CG + JacMax	**RFE**
91.12%	**91.97%**

Table 11.6 – State-of-the-art (SOTA) comparison on UKBench dataset (NS-Score).

N-S-Scores for state-of-the-art methods							
Qin *et al.* [267]	Zhang *et al.* [413]	Zheng *et al.* [424]	Bai *et al.* [16]	Xie *et al.* [371]	Lv *et al.* [210]	Liu *et al.* [203]	Pedronette *et al.* [241]
3.67	3.83	3.84	3.86	3.89	3.91	3.92	3.93
Bai *et al.* [20]	Liu *et al.* [159]	Valem *et al.* [328]	Bai *et al.* [17]	Valem *et al.* [329]	Valem *et al.* [327]	Chen *et al.* [50]	
3.93	3.93	3.93	3.94	3.94	3.95	3.96	

Proposed Approaches	
CG + JacMax	**RFE**
3.97	**3.97**

- **Discussion about Gains**

From the observed results, we can notice that the proposed approaches are comparable or better than state-of-the-art approaches in most cases. The best method in each scenario varies since each dataset and descriptor presents different aspects. An important attribute of the proposed approaches is their capacity to improve the input data by employing contextual similarity learning. Figure 11.1 presents the relative gains of RFE and JaccardMax for different datasets and descriptors. This demonstrates the capacity of contextual similarity learning to improve the results across multiple scenarios. The Holidays and UKBench datasets exhibited smaller gains because their descriptors already achieved higher results, making further enhancements more challenging compared to other datasets. Despite this, it is impressive that, despite the advancements in feature extraction, for different deep learning models from CNNs to Vision Transformers, the potential to obtain improved results was achieved across all cases.

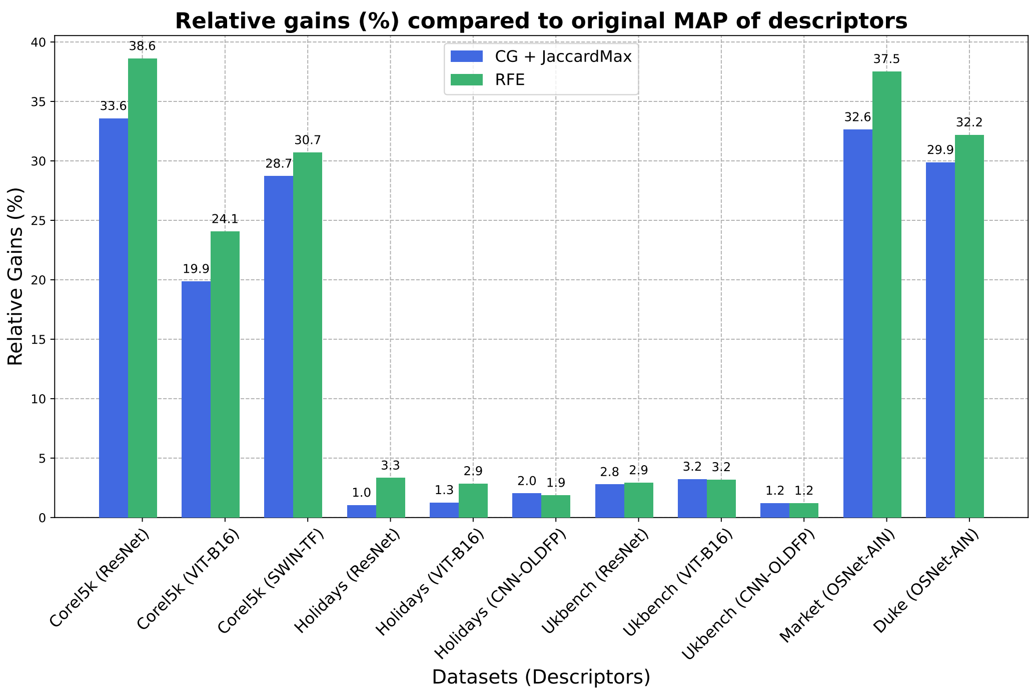

Figure 11.1 – RFE and JaccardMax relative gains (%) over MAP of descriptors.

11.1.3 Image Classification

Since both Manifold-GCN and RFE were employed for semi-supervised classification, Table 11.7 compares them to baselines, both traditional and recent, on Flowers and Corel5k datasets. The values achieved by Manifold-GCN are the highest in all the cases, and they are closely followed by RFE. These results reveal the high effectiveness of the proposed approaches that, besides the significant gains, are also comparable or superior to various methods in the literature.

Table 11.7 – Accuracy comparison (%) for baselines on Flowers and Corel5k datasets. We compared the proposed RFE and Manifold-GCN with semi-supervised classification baselines. The methods are compared with different input features. The results of our methods are highlighted with a gray background; the best results for each pair of features and dataset are marked in bold.

Method	Input	Flowers	Corel5k
CoMatch [169]	**Images**	82.55	*85.70*
kNN	**ResNet Features**	63.67	76.80
SVM [54]		80.54	88.73
OPF [8]		71.77	83.56
SL-Perceptron		75.44	83.56
ML-Perceptron		78.88	87.10
PseudoLabel+SGD [162]		82.69	89.76
LS+kNN [433]		73.49	83.98
LS+SVM [433, 54]		73.53	83.26
LS+OPF [433, 8]		72.66	82.32
LS+SL-Perceptron [433]		72.34	82.38
LS+ML-Perceptron [433]		73.03	82.53
GNN-LDS [90]		54.98	62.69
GNN-KNN-LDS [90]		79.32	88.94
WSEF [264]		85.12	91.68
RFE		84.95	91.54
Manifold-GCN		**85.88**	**93.08**
kNN	**SENet Features**	48.71	58.78
SVM [54]		73.30	85.89
OPF [8]		64.00	81.33
SL-Perceptron		71.84	82.28
ML-Perceptron		72.62	86.90
PseudoLabel+SGD [162]		76.87	89.85
LS+kNN [433]		58.05	72.16
LS+SVM [433, 54]		59.84	72.79
LS+OPF [433, 8]		59.25	72.20
LS+SL-Perceptron [433]		59.27	72.19
LS+ML-Perceptron [433]		59.39	72.24
GNN-LDS [90]		52.24	65.80
GNN-KNN-LDS [90]		73.69	89.95
WSEF [264]		76.16	89.74
RFE		77.56	92.20
Manifold-GCN		**78.82**	**92.79**

The CCL was also proposed and evaluated for classification but in supervised scenarios. Since it is the only method proposed in this category and it uses different datasets in the protocol, a direct comparison of it with other proposed methods is not feasible. Therefore, a discussion about its gains is presented. The experimental evaluation in Chapter 10 showed that the results are consistently better than those of SupCon, which CCL is based on, and SimCLR, another method commonly used as a baseline in this task. Figure 11.2 presents a plot that evinces the capacity of CCL to provide gains when compared to SupCon for three datasets and with higher values as the training set size decreases. The integration of contextual information within the contrastive loss significantly improved the results, as initially hypothesized, with gains up to 10.759%.

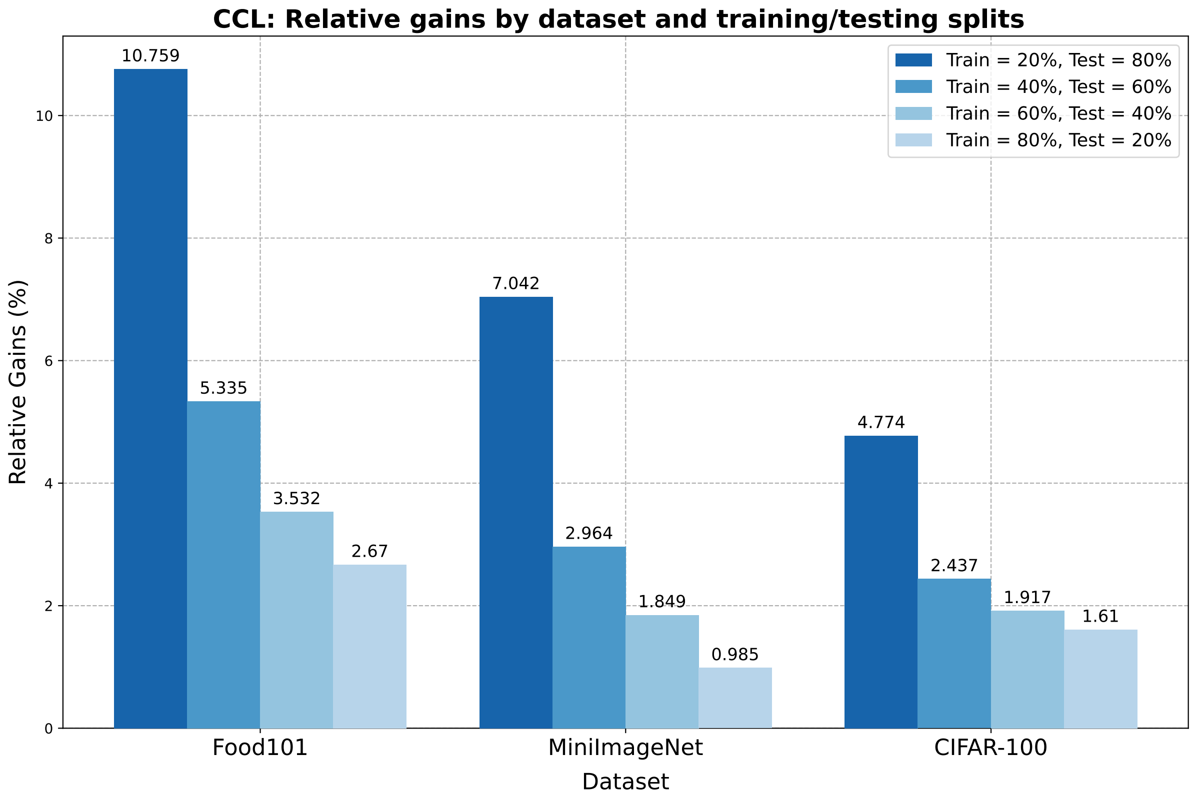

Figure 11.2 – Relative gains (%) obtained by CCL in comparison to SupCon for different train/test splits considering 100 training epochs.

Bibliography

[1] Agarwal, M. and Mostafa, J. (2011). Content-based image retrieval for alzheimer's disease detection. In *2011 9th International Workshop on Content-Based Multimedia Indexing (CBMI)*, pages 13–18.

[2] Albawi, S.; Mohammed, T. A.; and Al-Zawi, S. (2017). Understanding of a convolutional neural network. In *2017 International Conference on Engineering and Technology (ICET)*, pages 1–6.

[3] Ali, N.; Zafar, B.; Iqbal, M. K.; Sajid, M.; Younis, M. Y.; Dar, S. H.; Mahmood, M. T.; and Lee, I. H. (2019). Modeling global geometric spatial information for rotation invariant classification of satellite images. *PLoS One*, 14(7):e0219833.

[4] Alnissany, A. and Dayoub, Y. (2023). Modified centroid triplet loss for person re-identification. *Journal of Big Data*, 10(1):74.

[5] Alqasemi, F. A.; Alabbasi, H. Q.; Sabeha, F. G.; Alawadhi, A.; Kahlid, S.; and Zahary, A. (2019). Feature selection approach using knn supervised learning for content-based image retrieval. In *2019 First International Conference of Intelligent Computing and Engineering (ICOICE)*, pages 1–5.

[6] Alves, C. and Traina, A. J. M. (2022). Variational autoencoders for medical image retrieval. In *2022 International Conference on INnovations in Intelligent SysTems and Applications (INISTA)*, pages 1–6.

[7] Alvin, Y. H. Y. and Chakraborty, D. (2023). Approximate Maximum Rank Aggregation: Beyond the Worst-Case. In Bouyer, P. and Srinivasan, S., editors, *43rd IARCS Annual Conference on Foundations of Software Technology and Theoretical Computer Science (FSTTCS 2023)*, volume 284 of *Leibniz International Proceedings in Informatics (LIPIcs)*, pages 12:1–12:21, Dagstuhl, Germany. Schloss Dagstuhl – Leibniz-Zentrum für Informatik.

[8] Amorim, W. P.; Falcão, A. X.; and d. Carvalho, M. H. (2014). Semi-supervised pattern classification using optimum-path forest. In *2014 27th SIBGRAPI Conference on Graphics, Patterns and Images*, pages 111–118.

[9] An, L.; Chen, X.; Yang, S.; and Li, X. (2017). Person re-identification by multi-hypergraph fusion. *IEEE Transactions on Neural Networks and Learning Systems*, 28(11):2763–2774.

[10] Anand, A.; Leonhardt, J.; Rudra, K.; and Anand, A. (2022). Supervised contrastive learning approach for contextual ranking. In *Proceedings of the 2022 ACM SIGIR International Conference on Theory of Information Retrieval*, ICTIR ’22, page 61–71, New York, NY, USA. Association for Computing Machinery.

[11] Antelmi, A.; Cordasco, G.; Polato, M.; Scarano, V.; Spagnuolo, C.; and Yang, D. (2023). A survey on hypergraph representation learning. *ACM Comput. Surv.*, 56(1).

[12] Arandjelovic, R.; Gronat, P.; Torii, A.; Pajdla, T.; and Sivic, J. (2016). Netvlad: Cnn architecture for weakly supervised place recognition. In *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 5297–5307.

[13] Arica, N. and Vural, F. T. Y. (2003). BAS: a perceptual shape descriptor based on the beam angle statistics. *Pattern Recognition Letters*, 24(9-10):1627–1639.

[14] Awad, M. and Khanna, R. (2015). *Machine Learning*, pages 1–18. Apress, Berkeley, CA.

[15] Baeza-Yates, R. and Ribeiro-Neto, B. (2013). *Recuperação de Informação: Conceitos e Tecnologia das Máquinas de Busca*. Editora Bookman.

[16] Bai, S. and Bai, X. (2016). Sparse contextual activation for efficient visual re-ranking. *IEEE Trans. on Image Processing (TIP)*, 25(3):1056–1069.

[17] Bai, S.; Bai, X.; Tian, Q.; and Latecki, L. J. (2017). Regularized diffusion process for visual retrieval. In *Conf. on Artificial Intelligence (AAAI)*, pages 3967–3973.

[18] Bai, S.; Bai, X.; Tian, Q.; and Latecki, L. J. (2019). Regularized diffusion process on bidirectional context for object retrieval. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 41(5):1213–1226.

[19] Bai, S.; Zhang, F.; and Torr, P. H. (2021a). Hypergraph convolution and hypergraph attention. *Pattern Recognition*, 110:107637.

[20] Bai, S.; Zhou, Z.; Wang, J.; Bai, X.; Latecki, L. J.; and Tian, Q. (2017). Ensemble diffusion for retrieval. In *2017 IEEE International Conference on Computer Vision (ICCV)*, pages 774–783.

[21] Bai, X.; Bai, S.; and Wang, X. (2015). Beyond diffusion process: Neighbor set similarity for fast re-ranking. *Information Sciences*, 325:342 – 354.

[22] Bai, Z.; Wang, Z.; Wang, J.; Hu, D.; and Ding, E. (2021b). Unsupervised multi-source domain adaptation for person re-identification. In *2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 12909–12918.

[23] Bao, L.-N.; Wei, L.; Qiu, X.; gang Zhou, W.; Li, H.; and Tian, Q. (2023). Learning transferable pedestrian representation from multimodal information supervision. *ArXiv*, abs/2304.05554.

[24] Barz, B. and Denzler, J. (2021). Content-based image retrieval and the semantic gap in the deep learning era. In Del Bimbo, A.; Cucchiara, R.; Sclaroff, S.; Farinella, G. M.; Mei, T.; Bertini, M.; Escalante, H. J.; and Vezzani, R., editors, *Pattern Recognition. ICPR International Workshops and Challenges*, pages 245–260, Cham. Springer International Publishing.

[25] Bedagkar-Gala, A. and Shah, S. K. (2014). A survey of approaches and trends in person re-identification. *Image and Vision Computing*, 32(4):270 – 286.

[26] Berman, M.; Jégou, H.; Andrea, V.; Kokkinos, I.; and Douze, M. (2019). MultiGrain: a unified image embedding for classes and instances. *arXiv e-prints*.

[27] Berthelot, D.; Carlini, N.; Cubuk, E. D.; Kurakin, A.; Sohn, K.; Zhang, H.; and Raffel, C. (2020). Remixmatch: Semi-supervised learning with distribution matching and augmentation anchoring. In *International Conference on Learning Representations*.

[28] Berthelot, D.; Carlini, N.; Goodfellow, I. J.; Papernot, N.; Oliver, A.; and Raffel, C. (2019). Mixmatch: A holistic approach to semi-supervised learning. *CoRR*, abs/1905.02249.

[29] Bewley, A.; Ge, Z.; Ott, L.; Ramos, F.; and Upcroft, B. (2016). Simple online and realtime tracking. In *2016 IEEE International Conference on Image Processing (ICIP)*, pages 3464–3468.

[30] Bianchi, F. M.; Grattarola, D.; Livi, L.; and Alippi, C. (2021). Graph neural networks with convolutional arma filters. *IEEE TPAMI*, pages 1–1.

[31] Black, E. and Fredrikson, M. (2021). Leave-one-out unfairness. In *Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency*, FAccT '21, page 285–295, New York, NY, USA. Association for Computing Machinery.

[32] Bolme, D. S.; Beveridge, J. R.; Draper, B. A.; and Lui, Y. M. (2010). Visual object tracking using adaptive correlation filters. In *2010 IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, pages 2544–2550.

[33] Bossard, L.; Guillaumin, M.; and Van Gool, L. (2014). Food-101 – mining discriminative components with random forests. In Fleet, D.; Pajdla, T.; Schiele, B.; and Tuytelaars, T., editors, *Computer Vision – ECCV 2014*, pages 446–461, Cham. Springer International Publishing.

[34] Bretto, A. (2013). *Hypergraph Theory: An Introduction.* Springer International Publishing.

[35] Brodatz, P. (1966). *Textures: A Photographic Album for Artists and Designers.* Dover.

[36] Cai, D.; Zhang, C.; and He, X. (2010). Unsupervised feature selection for multi-cluster data. In *Proceedings of the 16th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, KDD '10, page 333–342, New York, NY, USA. Association for Computing Machinery.

[37] Camps, O.; Gou, M.; Hebble, T.; Karanam, S.; Lehmann, O.; Li, Y.; Radke, R. J.; Wu, Z.; and Xiong, F. (2017). From the lab to the real world: Re-identification in an airport camera network. *IEEE Transactions on Circuits and Systems for Video Technology*, 27(3):540–553.

[38] Chakraborty, D.; Das, S.; Khan, A.; and Subramanian, A. (2022). Fair rank aggregation. In *Advances in Neural Information Processing Systems 35 (NeurIPS 2022).*

[39] Chang, X.; Hospedales, T. M.; and Xiang, T. (2018). Multi-level factorisation net for person re-identification. In *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR).*

[40] Chatzichristofis, S. A. and Boutalis, Y. S. (2008a). Cedd: color and edge directivity descriptor: a compact descriptor for image indexing and retrieval. In *Proceedings of the 6th international conference on Computer vision systems*, ICVS'08, pages 312–322.

[41] Chatzichristofis, S. A. and Boutalis, Y. S. (2008b). Fcth: Fuzzy color and texture histogram - a low level feature for accurate image retrieval. In *WIAMIS*, pages 191–196.

[42] Chaudhuri, U.; Banerjee, B.; and Bhattacharya, A. (2019). Siamese graph convolutional network for content based remote sensing image retrieval. *Computer Vision and Image Understanding*, 184:22–30.

[43] Chen, H.; Lagadec, B.; and Bremond, F. (2021a). Ice: Inter-instance contrastive encoding for unsupervised person re-identification. In *2021 IEEE/CVF International Conference on Computer Vision (ICCV)*, pages 14940–14949.

[44] Chen, H.; Wang, Y.; Lagadec, B.; Dantcheva, A.; and Bremond, F. (2021b). Joint generative and contrastive learning for unsupervised person re-identification. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 2004–2013.

[45] Chen, S.-B.; Tian, X.-Z.; Ding, C. H. Q.; Luo, B.; Liu, Y.; Huang, H.; and Li, Q. (2020a). Graph convolutional network based on manifold similarity learning. *Cognitive Computation*, 12(6):1144.

[46] Chen, T. and Guestrin, C. (2016). Xgboost: A scalable tree boosting system. In *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, KDD '16, page 785–794, New York, NY, USA. Association for Computing Machinery.

[47] Chen, T.; Kornblith, S.; Norouzi, M.; and Hinton, G. (2020b). A simple framework for contrastive learning of visual representations. In *Proceedings of the 37th International Conference on Machine Learning*, ICML'20. JMLR.org.

[48] Chen, W.; Liu, Y.; Wang, W.; Bakker, E. M.; Georgiou, T.; Fieguth, P. W.; Liu, L.; and Lew, M. S. (2021c). Deep image retrieval: A survey. *CoRR*, abs/2101.11282.

[49] Chen, X. and He, K. (2021). Exploring simple siamese representation learning. In *2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 15745–15753.

[50] Chen, X. and Li, Y. (2020). Deep feature learning with manifold embedding for robust image retrieval. *Algorithms*, 13(12).

[51] Chen, Y.; Li, J.; Xiao, H.; Jin, X.; Yan, S.; and Feng, J. (2017). Dual path networks. In Guyon, I.; Luxburg, U. V.; Bengio, S.; Wallach, H.; Fergus, R.; Vishwanathan, S.; and Garnett, R., editors, *Advances in Neural Information Processing Systems 30*, pages 4467–4475. Curran Associates, Inc.

[52] Chollet, F. (2017). Xception: Deep learning with depthwise separable convolutions. In *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 1800–1807.

[53] Cieplinski, L. (2001). Mpeg-7 color descriptors and their applications. In Skarbek, W., editor, *Computer Analysis of Images and Patterns*, pages 11–20, Berlin, Heidelberg. Springer Berlin Heidelberg.

[54] Cortes, C. and Vapnik, V. (1995). Support-vector networks. *Mach. Learn.*, 20(3):273–297.

[55] Cristani, M. and Murino, V. (2018). Chapter 10 - person re-identification. In Chellappa, R. and Theodoridis, S., editors, *Academic Press Library in Signal Processing, Volume 6*, pages 365 – 394. Academic Press.

[56] Cronen-Townsend, S.; Zhou, Y.; and Croft, W. B. (2002). Predicting query performance. In *Proceedings of the 25th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*, SIGIR '02, page 299–306.

[57] Dabov, K.; Foi, A.; Katkovnik, V.; and Egiazarian, K. (2007). Image denoising by sparse 3-d transform-domain collaborative filtering. *IEEE Transactions on Image Processing*, 16(8):2080–2095.

[58] Dai, J.; Zhang, P.; Lu, H.; and Wang, H. (2018). Video person re-identification by temporal residual learning. *IEEE Transactions on Image Processing*, PP.

[59] Dalal, N. and Triggs, B. (2005). Histograms of oriented gradients for human detection. In *2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05)*, volume 1, pages 886–893 vol. 1.

[60] Datta, R.; Joshi, D.; Li, J.; and Wang, J. Z. (2008). Image retrieval: Ideas, influences, and trends of the new age. *ACM Computing Surveys*, 40(2):5:1–5:60.

[61] Datta, S.; Ganguly, D.; Greene, D.; and Mitra, M. (2022). Deep-qpp: A pairwise interaction-based deep learning model for supervised query performance prediction. In *Proceedings of the Fifteenth ACM International Conference on Web Search and Data Mining*, WSDM '22, page 201–209, New York, NY, USA. Association for Computing Machinery.

[62] De Almeida, L. B.; Pereira-Ferrero, V. H.; Valem, L. P.; Almeida, J.; and Pedronette, D. C. G. (2021). Representation learning for image retrieval through 3d cnn and manifold ranking. In *2021 34th SIBGRAPI Conference on Graphics, Patterns and Images (SIBGRAPI)*, pages 417–424.

[63] De Almeida, L. B.; Valem, L. P.; and Pedronette, D. C. G. (2022). Graph convolutional networks and manifold ranking for multimodal video retrieval. In *2022 IEEE International Conference on Image Processing (ICIP)*, pages 2811–2815.

[64] De Fernando, F. A.; Pedronette, D. C. G.; de Sousa, G. J.; Valem, L. P.; and Guilherme, I. R. (2020). Rade: A rank-based graph embedding approach. In *Proceedings of the 15th International Joint Conference on Computer Vision, Imaging and Computer Graphics Theory and Applications (VISIGRAPP 2020) - Volume 5: VISAPP*, pages 142–152. INSTICC, SciTePress.

[65] De Fernando, F. A.; Pedronette, D. C. G.; de Sousa, G. J.; Valem, L. P.; and Guilherme, I. R. (2022). Rade+: A semantic rank-based graph embedding algorithm. *International Journal of Information Management Data Insights*, 2(1):100078.

[66] De Oliveira, A. A.; Oakley, E.; Torres, R. d. S.; and Rocha, A. (2019). Relevance prediction in similarity-search systems using extreme value theory. *J. Vis. Commun. Image Represent.*, 60:236–249.

[67] De Sá, N. G.; Valem, L. P.; and Pedronette, D. C. G. (2021a). A multi-level rank correlation measure for image retrieval. In *VISIGRAPP (5: VISAPP)*, pages 370–378.

[68] De Sá, N. G.; Valem, L. P.; and Pedronette, D. C. G. (2021b). A multi-level rank correlation measure for image retrieval. In *Proceedings of the 16th International Joint Conference on Computer Vision, Imaging and Computer Graphics Theory and Applications (VISIGRAPP 2021) - Volume 5: VISAPP*, pages 370–378. INSTICC, SciTePress.

[69] Delvinioti, A.; Jégou, H.; Amsaleg, L.; and Houle, M. E. (2014). Image retrieval with reciprocal and shared nearest neighbors. In *Computer Vision Theory and Applications (VISAPP), 2014 International Conference on*, volume 2, pages 321–328.

[70] Deng, J.; Dong, W.; Socher, R.; Li, L.-J.; Li, K.; and Fei-Fei, L. (2009). ImageNet: A Large-Scale Hierarchical Image Database. In *CVPR09*.

[71] Deng, W.; Zheng, L.; Ye, Q.; Kang, G.; Yang, Y.; and Jiao, J. (2018). Image-image domain adaptation with preserved self-similarity and domain-dissimilarity for person re-identification. In *CVPR*.

[72] Dewil, V.; Barral, A.; Facciolo, G.; and Arias, P. (2022). Self-supervision versus synthetic datasets: which is the lesser evil in the context of video denoising? In *2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, pages 4896–4906, Los Alamitos, CA, USA. IEEE Computer Society.

[73] Dollár, P.; Appel, R.; Belongie, S.; and Perona, P. (2014). Fast feature pyramids for object detection. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 36(8):1532–1545.

[74] Donoser, M. and Bischof, H. (2013). Diffusion processes for retrieval revisited. In *2013 IEEE Conference on Computer Vision and Pattern Recognition*, pages 1320–1327.

[75] Dorogush, A. V.; Ershov, V.; and Gulin, A. (2018). Catboost: gradient boosting with categorical features support.

[76] Dos Santos, C. F. G.; de Souza Oliveira, D.; Passos, L. A.; Pires, R. G.; Santos, D. F. S.; Valem, L. P.; Moreira, T. P.; Santana, M. C. S.; Roder, M.; Papa, J. P.; and Colombo, D. (2022). Gait recognition based on deep learning: A survey. *ACM Comput. Surv.*, 55(2).

[77] Dosovitskiy, A.; Beyer, L.; Kolesnikov, A.; Weissenborn, D.; Zhai, X.; Unterthiner, T.; Dehghani, M.; Minderer, M.; Heigold, G.; Gelly, S.; Uszkoreit, J.; and Houlsby, N. (2021). An image is worth 16x16 words: Transformers for image recognition at scale. In *International Conference on Learning Representations*.

[78] Dou, Z.; Cui, H.; Zhang, L.; and Wang, B. (2020). Learning global and local consistent representations for unsupervised image retrieval via deep graph diffusion networks. *arXiv preprint arXiv:2001.01284*.

[79] Drucker, H.; Burges, C. J. C.; Kaufman, L.; Smola, A.; and Vapnik, V. (1996). Support vector regression machines. In Mozer, M.; Jordan, M.; and Petsche, T., editors, *Advances in Neural Information Processing Systems*, volume 9. MIT Press.

[80] Dubey, S. R. (2022). A decade survey of content based image retrieval using deep learning. *IEEE Transactions on Circuits and Systems for Video Technology*, 32(5):2687–2704.

[81] Duncanson, K. A.; Thwaites, S.; Booth, D.; Hanly, G.; Robertson, W. S. P.; Abbasnejad, E.; and Thewlis, D. (2023). Deep metric learning for scalable gait-based person re-identification using force platform data. *Sensors*, 23(7).

[82] Dwibedi, D.; Aytar, Y.; Tompson, J.; Sermanet, P.; and Zisserman, A. (2021). With a little help from my friends: Nearest-neighbor contrastive learning of visual representations. *2021 IEEE/CVF International Conference on Computer Vision (ICCV)*, pages 9568–9577.

[83] Dwork, C.; Kumar, R.; Naor, M.; and Sivakumar, D. (2001). Rank aggregation methods for the web. In *Proceedings of the 10th International Conference on World Wide Web*, WWW ’01, page 613–622, New York, NY, USA. Association for Computing Machinery.

[84] Elsayed, G.; Krishnan, D.; Mobahi, H.; Regan, K.; and Bengio, S. (2018). Large margin deep networks for classification. In *Advances in neural information processing systems*, pages 842–852.

[85] Engelen, J. E. V. and Hoos, H. H. (2019). A survey on semi-supervised learning. *Machine Learning*, 109:373–440.

[86] Escolano, F.; Hancock, E. R.; Lozano, M. A.; and Curado, M. (2017). The mutual information between graphs. *Pattern Recognition Letters*, 87:12 – 19.

[87] Fagin, R.; Kumar, R.; and Sivakumar, D. (2003). Comparing top k lists. In *Proceedings of the Fourteenth Annual ACM-SIAM Symposium on Discrete Algorithms*, SODA ’03, page 28–36, USA. Society for Industrial and Applied Mathematics.

[88] Figueira, D.; Bazzani, L.; Minh, H. Q.; Cristani, M.; Bernardino, A.; and Murino, V. (2013). Semi-supervised multi-feature learning for person re-identification. In *2013 10th IEEE International Conference on Advanced Video and Signal Based Surveillance*, pages 111–116.

[89] Fogel, I. and Sagi, D. (1989). Gabor filters as texture discriminator. *Biological Cybernetics*, 61(2):103–113.

[90] Franceschi, L.; Niepert, M.; Pontil, M.; and He, X. (2019). Learning discrete structures for graph neural networks. In *Proceedings of the 36th International Conference on Machine Learning*.

[91] Fredriksson, T.; Mattos, D. I.; Bosch, J.; and Olsson, H. H. (2020). Data labeling: An empirical investigation into industrial challenges and mitigation strategies. In Morisio, M.; Torchiano, M.; and Jedlitschka, A., editors, *Product-Focused Software Process Improvement*, pages 202–216, Cham. Springer International Publishing.

[92] Friedman, J. H. (2001). Greedy function approximation: A gradient boosting machine. *The Annals of Statistics*, 29(5):1189 – 1232.

[93] Fu, Z.; Li, Y.; Mao, Z.; Wang, Q.; and Zhang, Y. (2021). Deep metric learning with self-supervised ranking. *Proceedings of the AAAI Conference on Artificial Intelligence*, 35(2):1370–1378.

[94] García, J.; Martinel, N.; Gardel, A.; Bravo, I.; Foresti, G. L.; and Micheloni, C. (2016). Modeling feature distances by orientation driven classifiers for person re-identification. *Journal of Visual Communication and Image Representation*, 38:115 – 129.

[95] García, J.; Martinel, N.; Gardel, A.; Bravo, I.; Foresti, G. L.; and Micheloni, C. (2017). Discriminant context information analysis for post-ranking person re-identification. *IEEE Transactions on Image Processing*, 26(4):1650–1665.

[96] García, J.; Martinel, N.; Micheloni, C.; and Gardel, A. (2015). Person re-identification ranking optimisation by discriminant context information analysis. In *2015 IEEE International Conference on Computer Vision (ICCV)*, pages 1305–1313.

[97] Ge, W.; Pan, C.; Wu, A.; Zheng, H.; and Zheng, W.-S. (2021). *Cross-Camera Feature Prediction for Intra-Camera Supervised Person Re-Identification across Distant Scenes*, page 3644–3653. Association for Computing Machinery, New York, NY, USA.

[98] Geusebroek, J.-M.; Burghouts, G. J.; and Smeulders, A. W. M. (2005). The amsterdam library of object images. *International Journal of Computer Vision*, 61(1):103–112.

[99] Gionis, A.; Indyk, P.; and Motwani, R. (1999). Similarity search in high dimensions via hashing. In *Proceedings of the 25th International Conference on Very Large Data Bases*, VLDB '99, page 518–529.

[100] Giovanni, F. D.; Rowbottom, J.; Chamberlain, B. P.; Markovich, T.; and Bronstein, M. M. (2022). Graph neural networks as gradient flows: understanding graph convolutions via energy. *arXiv preprint arXiv:2206.10991*.

[101] Gong, C.; Wang, D.; and Liu, Q. (2021). Alphamatch: Improving consistency for semi-supervised learning with alpha-divergence. In *2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 13678–13687.

[102] Gopalakrishnan, S.; Sridharan, S.; Nayak, S. R.; Nayak, J.; and Venkataraman, S. (2022). Central hubs prediction for bio networks by directed hypergraph - ga with validation to covid-19 ppi. *Pattern Recognition Letters*, 153:246–253.

[103] Gopalan, R.; Turaga, P.; and Chellappa, R. (2010). Articulation-invariant representation of non-planar shapes. In *ECCV*, volume 3, pages 286–299.

[104] Gordo, A.; Almazán, J.; Revaud, J.; and Larlus, D. (2017). End-to-end learning of deep visual representations for image retrieval. *International Journal of Computer Vision*, 124.

[105] Grabner, H.; Grabner, M.; and Bischof, H. (2006). Real-time tracking via on-line boosting. In Chantler, M. J.; Fisher, R. B.; and Trucco, E., editors, *BMVC*, pages 47–56. British Machine Vision Association.

[106] Gray, D. and Tao, H. (2008). Viewpoint invariant pedestrian recognition with an ensemble of localized features. In Forsyth, D.; Torr, P.; and Zisserman, A., editors, *Computer Vision – ECCV 2008*, pages 262–275, Berlin, Heidelberg. Springer Berlin Heidelberg.

[107] Gui, J.; Chen, T.; Zhang, J.; Cao, Q.; Sun, Z.; Luo, H.; and Tao, D. (2023). A survey on self-supervised learning: Algorithms, applications, and future trends.

[108] Guo, R.; Li, C.; Li, Y.; and Lin, J. (2018). Density-adaptive kernel based re-ranking for person re-identification. In *2018 24th International Conference on Pattern Recognition (ICPR)*, pages 982–987.

[109] Hammoudeh, Z. and Lowd, D. (2024). Training data influence analysis and estimation: a survey. *Machine Learning*, 113(5):2351–2403.

[110] He, K.; Zhang, X.; Ren, S.; and Sun, J. (2016). Deep residual learning for image recognition. In *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 770–778.

[111] He, S.; Luo, H.; Wang, P.; Wang, F.; Li, H.; and Jiang, W. (2021). Transreid: Transformer-based object re-identification. In *ICCV*, pages 15013–15022.

[112] He, X.; Cai, D.; and Niyogi, P. (2005). Laplacian score for feature selection. In *Proceedings of the 18th International Conference on Neural Information Processing Systems*, NIPS'05, pages 507–514, Cambridge, MA, USA. MIT Press.

[113] Henriques, J. F.; Caseiro, R.; Martins, P.; and Batista, J. (2015). High-speed tracking with kernelized correlation filters. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 37(3):583–596.

[114] Hermans, A.; Beyer, L.; and Leibe, B. (2017). In Defense of the Triplet Loss for Person Re-Identification. *arXiv preprint arXiv:1703.07737*.

[115] Hoi, S. C.; Liu, W.; and Chang, S.-F. (2010). Semi-supervised distance metric learning for collaborative image retrieval and clustering. *ACM Transactions on Multimedia Computing and Communication Applications*, 6(3):18:1–18:26.

[116] Hoyer, P. O. (2004). Non-negative matrix factorization with sparseness constraints. *J. Mach. Learn. Res.*, 5:1457–1469.

[117] Hu, J.; Shen, L.; and Sun, G. (2018). Squeeze-and-excitation networks. In *2018 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*.

[118] Huang, H.; Yang, W.; Chen, X.; Zhao, X.; Huang, K.; Lin, J.; Huang, G.; and Du, D. (2018). Eanet: Enhancing alignment for cross-domain person re-identification. *arXiv preprint arXiv:1812.11369*.

[119] Huang, J.; Kumar, S. R.; Mitra, M.; Zhu, W.-J.; and Zabih, R. (1997). Image indexing using color correlograms. In *CVPR*, pages 762–768.

[120] Huang, Y.; Liu, Q.; Zhang, S.; and Metaxas, D. N. (2010). Image retrieval via probabilistic hypergraph ranking. In *IEEE Conference on Conference on Computer Vision and Pattern Recognition (CVPR'10)*, pages 3376–3383.

[121] Huang, Y.; Peng, P.; Jin, Y.; Xing, J.; Lang, C.; and Feng, S. (2019). Domain adaptive attention model for unsupervised cross-domain person re-identification. *CoRR*, abs/1905.10529.

[122] Ioffe, S. and Szegedy, C. (2015). Batch normalization: Accelerating deep network training by reducing internal covariate shift. In *Proceedings of the 32Nd International Conference on International Conference on Machine Learning - Volume 37*, ICML'15, pages 448–456. JMLR.org.

[123] Iscen, A.; Avrithis, Y.; Tolias, G.; Furon, T.; and Chum, O. (2018a). Fast spectral ranking for similarity search. In *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 7632–7641.

[124] Iscen, A.; Tolias, G.; Avrithis, Y.; and Chum, O. (2018b). Mining on manifolds: Metric learning without labels. In *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 7642–7651.

[125] Iscen, A.; Tolias, G.; Avrithis, Y.; Furon, T.; and Chum, O. (2017). Efficient diffusion on region manifolds: Recovering small objects with compact cnn representations. In *CVPR*.

[126] Izenman, A. J. (2008). *Linear Discriminant Analysis*, pages 237–280. Springer New York, New York, NY.

[127] Jegou, H.; Douze, M.; and Schmid, C. (2008). Hamming embedding and weak geometric consistency for large scale image search. In *European Conference on Computer Vision*, ECCV '08, pages 304–317.

[128] Jegou, H.; Schmid, C.; Harzallah, H.; and Verbeek, J. (2010). Accurate image search using the contextual dissimilarity measure. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 32(1):2–11.

[129] Jensen, J. L. W. V. (1906). Sur les fonctions convexes et les inégalités entre les valeurs moyennes. *Acta Mathematica*, 30(1):175–193.

[130] Jia, Q. and Tian, X. (2015). Query difficulty estimation via relevance prediction for image retrieval. *Signal Processing*, 110:232–243.

[131] Jia, Q.; Tian, X.; and Mei, T. (2014). Query difficulty estimation via pseudo relevance feedback for image search. In *2014 IEEE International Conference on Multimedia and Expo (ICME)*, pages 1–6.

[132] Jia, X.; Zhong, X.; Ye, M.; Liu, W.; and Huang, W. (2022). Complementary data augmentation for cloth-changing person re-identification. *IEEE Transactions on Image Processing*, 31:4227–4239.

[133] Jiang, J.; Wang, B.; and Tu, Z. (2011). Unsupervised metric learning by self-smoothing operator. In *2011 International Conference on Computer Vision*, pages 794–801.

[134] Jin, X.; Lan, C.; Zeng, W.; Chen, Z.; and Zhang, L. (2020). Style normalization and restitution for generalizable person re-identification. In *2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 3140–3149.

[135] Ju, W.; Yi, S.; Wang, Y.; Xiao, Z.; Mao, Z.; Li, H.; Gu, Y.; Qin, Y.; Yin, N.; Wang, S.; Liu, X.; Luo, X.; Yu, P. S.; and Zhang, M. (2024). A survey of graph neural networks in real world: Imbalance, noise, privacy and ood challenges. *arXiv preprint arXiv:2403.04468*.

[136] Juan, D.; Lu, C.; Li, Z.; Peng, F.; Timofeev, A.; Chen, Y.; Gao, Y.; Duerig, T.; Tomkins, A.; and Ravi, S. (2019). Graph-rise: Graph-regularized image semantic embedding. *CoRR*, abs/1902.10814.

[137] Karanam, S.; Gou, M.; Wu, Z.; Rates-Borras, A.; Camps, O.; and Radke, R. J. (2019). A systematic evaluation and benchmark for person re-identification: Features, metrics, and datasets. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 41(3):523–536.

[138] Kawai, V. S.; Valem, L. P.; Baldassin, A.; Borin, E.; Pedronette, D. C. G.; and Latecki, L. J. (2024). Rank-based hashing for effective and efficient nearest neighbor search for image retrieval. *ACM Transactions on Multimedia Computing, Communications, and Applications*.

[139] Ke, G.; Meng, Q.; Finley, T.; Wang, T.; Chen, W.; Ma, W.; Ye, Q.; and Liu, T.-Y. (2017). Lightgbm: A highly efficient gradient boosting decision tree. In *Proceedings of the 31st International Conference on Neural Information Processing Systems*, NIPS'17, page 3149–3157, Red Hook, NY, USA. Curran Associates Inc.

[140] Khasanova, A. M. and Pasechnik, M. O. (2021). Social media analysis with machine learning. In *2021 IEEE Conference of Russian Young Researchers in Electrical and Electronic Engineering (ElConRus)*, pages 32–35.

[141] Khemani, B.; Patil, S.; Kotecha, K.; and Tanwar, S. (2024). A review of graph neural networks: concepts, architectures, techniques, challenges, datasets, applications, and future directions. *Journal of Big Data*, 11(1):18.

[142] Khosla, A.; Jayadevaprakash, N.; Yao, B.; and Fei-Fei, L. (2011). Novel dataset for fine-grained image categorization. In *Workshop on Fine-Grained Visual Categorization, CVPR*.

[143] Khosla, P.; Teterwak, P.; Wang, C.; Sarna, A.; Tian, Y.; Isola, P.; Maschinot, A.; Liu, C.; and Krishnan, D. (2020). Supervised contrastive learning. In Larochelle, H.; Ranzato, M.; Hadsell, R.; Balcan, M.; and Lin, H., editors, *Advances in Neural Information Processing Systems*, volume 33, pages 18661–18673. Curran Associates, Inc.

[144] Khraimeche, Y.; Bilodeau, G.-A.; Steele, D.; and Mahadik, H. (2020). Unsupervised disentanglement gan for domain adaptive person re-identification. *ArXiv*, abs/2007.15560.

[145] Kim, B.; Choo, J.; Kwon, Y.; Joe, S.; Min, S.; and Gwon, Y. (2021). Selfmatch: Combining contrastive self-supervision and consistency for semi-supervised learning. *CoRR*, abs/2101.06480.

[146] Kipf, T. N. and Welling, M. (2017). Semi-supervised classification with graph convolutional networks. In *5th International Conference on Learning Representations,*

ICLR 2017, Toulon, France, April 24-26, 2017, Conference Track Proceedings. OpenReview.net.

[147] Klicpera, J.; Bojchevski, A.; and Günnemann, S. (2019). Combining neural networks with personalized pagerank for classification on graphs. In *International Conference on Learning Representations*.

[148] Kovalev, V. and Volmer, S. (1998). Color co-occurence descriptors for querying-by-example. In *International Conference on Multimedia Modeling (ICMM)*, page 32.

[149] Kreiss, S.; Bertoni, L.; and Alahi, A. (2019). Pifpaf: Composite fields for human pose estimation. *CoRR*, abs/1903.06593.

[150] Krizhevsky, A.; Nair, V.; and Hinton, G. E. (2009). Learning multiple layers of features from tiny images. Technical report, University of Toronto. Technical Report TR-2009.

[151] Krizhevsky, A.; Sutskever, I.; and Hinton, G. E. (2012). Imagenet classification with deep convolutional neural networks. In Pereira, F.; Burges, C. J. C.; Bottou, L.; and Weinberger, K. Q., editors, *Advances in Neural Information Processing Systems 25*, pages 1097–1105. Curran Associates, Inc.

[152] Krull, A.; Buchholz, T.; and Jug, F. (2019). Noise2void - learning denoising from single noisy images. In *2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 2124–2132.

[153] Kumar, D.; Siva, P.; Marchwica, P.; and Wong, A. (2019). Fairest of them all: Establishing a strong baseline for cross-domain person reid. *CoRR*, abs/1907.12016.

[154] Kuncheva, L. I. (2004). *Combining Pattern Classifiers: Methods and Algorithms*. Wiley.

[155] Kurland, O. (2014). The cluster hypothesis in information retrieval. In de Rijke, M.; Kenter, T.; de Vries, A. P.; Zhai, C.; de Jong, F.; Radinsky, K.; and Hofmann, K., editors, *Advances in Information Retrieval*, pages 823–826, Cham. Springer International Publishing.

[156] Köstinger, M.; Hirzer, M.; Wohlhart, P.; Roth, P. M.; and Bischof, H. (2012). Large scale metric learning from equivalence constraints. In *2012 IEEE Conference on Computer Vision and Pattern Recognition*, pages 2288–2295.

[157] Laine, S. and Aila, T. (2017). Temporal ensembling for semi-supervised learning. In *International Conference on Learning Representations*.

[158] Laine, S.; Karras, T.; Lehtinen, J.; and Aila, T. (2019). High-quality self-supervised deep image denoising. In Wallach, H.; Larochelle, H.; Beygelzimer, A.; d'Alché-Buc, F.;

Fox, E.; and Garnett, R., editors, *Advances in Neural Information Processing Systems 32*, pages 6970–6980. Curran Associates, Inc.

[159] Lao, G.; Liu, S.; Tan, C.; Wang, Y.; Li, G.; Xu, L.; Feng, L.; and Wang, F. (2021). Three degree binary graph and shortest edge clustering for re-ranking in multi-feature image retrieval. *Journal of Visual Communication and Image Representation*, 80:103282.

[160] Lary, D. J.; Alavi, A. H.; Gandomi, A. H.; and Walker, A. L. (2016). Machine learning in geosciences and remote sensing. *Geoscience Frontiers*, 7(1):3–10. Special Issue: Progress of Machine Learning in Geosciences.

[161] Latecki, L. J.; Lakamper, R.; and Eckhardt, U. (2000). Shape descriptors for non-rigid shapes with a single closed contour. In *CVPR*, pages 424–429.

[162] Lee, D.-H. (2013). Pseudo-label: The simple and efficient semi-supervised learning method for deep neural networks. In *Workshop on challenges in representation learning, ICML*, volume 3.

[163] Lee, K.; Jang, I.-S.; Kim, K.-J.; and Kim, P.-K. (2022). Reet: Region-enhanced transformer for person re-identification. In *2022 18th IEEE International Conference on Advanced Video and Signal Based Surveillance (AVSS)*, pages 1–8.

[164] Lehtinen, J.; Munkberg, J.; Hasselgren, J.; Laine, S.; Karras, T.; Aittala, M.; and Aila, T. (2018). Noise2Noise: Learning image restoration without clean data. In Dy, J. and Krause, A., editors, *Proceedings of the 35th International Conference on Machine Learning*, volume 80 of *Proceedings of Machine Learning Research*, pages 2965–2974. PMLR.

[165] Leng, Q.; Hu, R.; Liang, C.; Wang, Y.; and Chen, J. (2014). Person re-identification with content and context re-ranking. *Multimedia Tools and Applications*, 74.

[166] Leticio, G.; Valem, L. P.; Lopes, L. T.; and Pedronette, D. C. G. (2023). pyudlf: A python framework for unsupervised distance learning tasks. In *Proceedings of the 31st ACM International Conference on Multimedia*, MM '23, page 9680–9684, New York, NY, USA. Association for Computing Machinery.

[167] Leticio, G. R.; Kawai, V. S.; Valem, L. P.; Pedronette, D. C. G.; and da S. Torres, R. (2024). Manifold information through neighbor embedding projection for image retrieval. *Pattern Recognition Letters*, 183:17–25.

[168] Li, D.; Li, D.; Zhang, Z.; Wang, L.; and Tan, T. (2019a). Unsupervised cross-domain person re-identification: A new framework. In *2019 IEEE International Conference on Image Processing (ICIP)*, pages 1222–1226.

[169] Li, J.; Xiong, C.; and Hoi, S. C. H. (2021a). Comatch: Semi-supervised learning with contrastive graph regularization. In *2021 IEEE/CVF International Conference on Computer Vision (ICCV)*, pages 9455–9464.

[170] Li, M.; Zhu, X.; and Gong, S. (2018a). Unsupervised person re-identification by deep learning tracklet association. In *Proceedings of the European Conference on Computer Vision (ECCV)*, pages 737–753.

[171] Li, M.; Zhu, X.; and Gong, S. (2019b). Unsupervised tracklet person re-identification. *IEEE Transactions on Pattern Analysis and Machine Intelligence.*

[172] Li, Q.; Han, Z.; and Wu, X.-m. (2018b). Deeper insights into graph convolutional networks for semi-supervised learning. *Proceedings of the AAAI Conference on Artificial Intelligence*, 32(1).

[173] Li, S.; Gao, C.; Yu, H.; and Zhang, J. (2016). Person re-identification via person dpm based partition. In *2016 23rd International Conference on Pattern Recognition (ICPR)*, pages 3856–3861.

[174] Li, W.; Wu, Y.; Mukunoki, M.; and Minoh, M. (2012). Common-near-neighbor analysis for person re-identification. In *2012 19th IEEE International Conference on Image Processing*, pages 1621–1624.

[175] Li, W.; Zhao, R.; and Wang, X. (2012a). Human reidentification with transferred metric learning. In *ACCV*.

[176] Li, W.; Zhao, R.; Xiao, T.; and Wang, X. (2014). Deepreid: Deep filter pairing neural network for person re-identification. In *2014 IEEE Conference on Computer Vision and Pattern Recognition*, pages 152–159.

[177] Li, W.; Zhu, X.; and Gong, S. (2018c). Harmonious attention network for person re-identification. In *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*.

[178] Li, X.; Larson, M.; and Hanjalic, A. (2015). Pairwise geometric matching for large-scale object retrieval. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR'2015)*, pages 5153–5161.

[179] Li, Y.; Geng, B.; Yang, L.; Xu, C.; and Bian, W. (2012b). Query difficulty estimation for image retrieval. *Neurocomputing*, 95:48–53.

[180] Li, Y.; Wu, Z.; Karanam, S.; and Radke, R. J. (2014). Real-world re-identification in an airport camera network. In *Proceedings of the International Conference on Distributed Smart Cameras*, ICDSC '14, New York, NY, USA. Association for Computing Machinery.

[181] Li, Y.-J.; Yang, F.-E.; Liu, Y.-C.; Yeh, Y.-Y.; Du, X.; and Frank Wang, Y.-C. (2018d). Adaptation and re-identification network: An unsupervised deep transfer learning approach to person re-identification. In *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR) Workshops*.

[182] Li, Z.; Yang, Y.; Liu, J.; Zhou, X.; and Lu, H. (2021b). Unsupervised feature selection using nonnegative spectral analysis. *Proceedings of the AAAI Conference on Artificial Intelligence*, 26(1):1026–1032.

[183] Liao, C.; Tsiligkaridis, T.; and Kulis, B. (2023). Supervised metric learning to rank for retrieval via contextual similarity optimization. In *Proceedings of the 40th International Conference on Machine Learning*, ICML'23. JMLR.org.

[184] Liao, S.; Hu, Y.; Xiangyu Zhu; and Li, S. Z. (2015). Person re-identification by local maximal occurrence representation and metric learning. In *2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 2197–2206.

[185] Liao, S. and Shao, L. (2022a). Graph sampling based deep metric learning for generalizable person re-identification. In *2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 7349–7358.

[186] Liao, S. and Shao, L. (2022b). Graph sampling based deep metric learning for generalizable person re-identification. In *2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 7349–7358.

[187] Liao, S.; Zhao, G.; Kellokumpu, V.; Pietikäinen, M.; and Li, S. Z. (2010). Modeling pixel process with scale invariant local patterns for background subtraction in complex scenes. In *2010 IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, pages 1301–1306.

[188] Lin, Y.; Wu, Y.; Yan, C.; Xu, M.; and Yang, Y. (2020a). Unsupervised person re-identification via cross-camera similarity exploration. *IEEE Transactions on Image Processing*.

[189] Lin, Y.; Xie, L.; Wu, Y.; Yan, C.; and Tian, Q. (2020b). Unsupervised person re-identification via softened similarity learning. In *2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 3387–3396.

[190] Ling, H. and Jacobs, D. W. (2007). Shape classification using the inner-distance. *IEEE TPAMI*, 29(2):286–299.

[191] Ling, H.; Yang, X.; and Latecki, L. J. (2010). Balancing deformability and discriminability for shape matching. In *ECCV*, volume 3, pages 411–424.

[192] Lisanti, G.; Masi, I.; Bagdanov, A. D.; and Bimbo, A. D. (2015). Person re-identification by iterative re-weighted sparse ranking. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 37(8):1629–1642.

[193] Liu, C.; Yu, G.; Volkovs, M.; Chang, C.; Rai, H.; Ma, J.; and Gorti, S. K. (2019a). Guided similarity separation for image retrieval. In Wallach, H.; Larochelle, H.; Beygelzimer, A.; d'Alché-Buc, F.; Fox, E.; and Garnett, R., editors, *Advances in Neural Information Processing Systems*, volume 32. Curran Associates, Inc.

[194] Liu, G.-H. and Yang, J.-Y. (2013). Content-based image retrieval using color difference histogram. *Pattern Recognition*, 46(1):188 – 198.

[195] Liu, H.; Cheng, J.; Wang, S.; and Wang, W. (2019b). Attention: A big surprise for cross-domain person re-identification. *CoRR*, abs/1905.12830.

[196] Liu, H.; Feng, J.; Qi, M.; Jiang, J.; and Yan, S. (2016a). End-to-end comparative attention networks for person re-identification. *CoRR*, abs/1606.04404.

[197] Liu, J.; Zha, Z.-J.; Chen, D.; Hong, R.; and Wang, M. (2019c). Adaptive transfer network for cross-domain person re-identification. In *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*.

[198] Liu, S. and Deng, W. (2015). Very deep convolutional neural network based image classification using small training sample size. In *2015 3rd IAPR Asian Conference on Pattern Recognition (ACPR)*, pages 730–734.

[199] Liu, W.; Wen, Y.; Yu, Z.; and Yang, M. (2016b). Large-margin softmax loss for convolutional neural networks. In *Proceedings of the International Conference on Machine Learning (ICML)*, volume 2.

[200] Liu, X.; Song, M.; Tao, D.; Zhou, X.; Chen, C.; and Bu, J. (2014). Semi-supervised coupled dictionary learning for person re-identification. In *2014 IEEE Conference on Computer Vision and Pattern Recognition*, pages 3550–3557.

[201] Liu, Y.; Ge, H.; Wang, Z.; Hou, Y.; and Zhao, M. (2024). Discriminative identity-feature exploring and differential aware learning for unsupervised person re-identification. *IEEE Transactions on Multimedia*, 26:623–636.

[202] Liu, Z.; Lin, Y.; Cao, Y.; Hu, H.; Wei, Y.; Zhang, Z.; Lin, S.; and Guo, B. (2021). Swin transformer: Hierarchical vision transformer using shifted windows. *CoRR*, abs/2103.14030.

[203] Liu, Z.; Wang, S.; Zheng, L.; and Tian, Q. (2017). Robust imagegraph: Rank-level feature fusion for image search. *IEEE Transactions on Image Processing*, 26(7):3128–3141.

[204] Lopes, L. T.; Valem, L. P.; Pedronette, D. C. G.; Guilherme, I. R.; Papa., J. P.; Santana, M. C. S.; and Colombo, D. (2020). Manifold learning-based clustering approach applied to anomaly detection in surveillance videos. In *Proceedings of the 15th International Joint Conference on Computer Vision, Imaging and Computer Graphics Theory and Applications - Volume 5: VISAPP,*, pages 404–412. INSTICC, SciTePress.

[205] Lowe, D. (1999). Object recognition from local scale-invariant features. In *ICCV*, pages 1150–1157.

[206] Loy, C. C.; Xiang, T.; and Gong, S. (2009). Multi-camera activity correlation analysis. In *2009 IEEE Conference on Computer Vision and Pattern Recognition*, pages 1988–1995.

[207] Lukežič, A.; Vojíř, T.; Čehovin Zajc, L.; Matas, J.; and Kristan, M. (2018). Discriminative correlation filter with channel and spatial reliability. *International Journal of Computer Vision*, 126.

[208] Luo, X.; Ju, W.; Gu, Y.; Mao, Z.; Liu, L.; Yuan, Y.; and Zhang, M. (2023). Self-supervised graph-level representation learning with adversarial contrastive learning. *ACM Trans. Knowl. Discov. Data*, 18(2).

[209] Lux, M. (2011). Content based image retrieval with lire. In *Proceedings of the 19th ACM International Conference on Multimedia*, MM '11, pages 735–738, New York, NY, USA. ACM.

[210] Lv, Y.; Zhou, W.; Tian, Q.; Sun, S.; and Li, H. (2018). Retrieval oriented deep feature learning with complementary supervision mining. *IEEE Transactions on Image Processing*, 27(10):4945–4957.

[211] Ma, A. J. and Li, P. (2015). Query based adaptive re-ranking for person re-identification. In Cremers, D.; Reid, I.; Saito, H.; and Yang, M.-H., editors, *Computer Vision – ACCV 2014*, pages 397–412, Cham. Springer International Publishing.

[212] Ma, B.; Su, Y.; and Jurie, F. (2012). Local descriptors encoded by fisher vectors for person re-identification. In Fusiello, A.; Murino, V.; and Cucchiara, R., editors, *Computer Vision – ECCV 2012. Workshops and Demonstrations*, pages 413–422, Berlin, Heidelberg. Springer Berlin Heidelberg.

[213] Ma, B.; Su, Y.; and Jurie, F. (2014). Covariance descriptor based on bio-inspired features for person re-identification and face verification. *Image and Vision Computing*, 32(6):379 – 390.

[214] Maaten, L. V. D. and Hinton, G. E. (2008). Visualizing data using t-sne. *Journal of Machine Learning Research*, 9:2579–2605.

[215] Manjunath, B. S.; Ohm, J. R.; Vasudevan, V. V.; and Yamada, A. (2001). Color and texture descriptors. *IEEE Transactions on Circuits and Systems for Video Technology*, 11(6):703–715.

[216] Martinel, N.; Micheloni, C.; and Foresti, G. L. (2015). Kernelized saliency-based person re-identification through multiple metric learning. *IEEE Transactions on Image Processing*, 24(12):5645–5658.

[217] Matsukawa, T.; Okabe, T.; Suzuki, E.; and Sato, Y. (2016). Hierarchical gaussian descriptor for person re-identification. In *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 1363–1372.

[218] Meishvili, G.; Djelouah, A.; Hattori, S.; and Schroers, C. (2022). Contrastive learning for controllable blind video restoration. In *33rd British Machine Vision Conference 2022, BMVC 2022, London, UK, November 21-24, 2022*. BMVA Press.

[219] Meng, J.; Wu, S.; and Zheng, W.-S. (2019). Weakly supervised person re-identification. In *2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 760–769.

[220] Miyato, T.; ichi Maeda, S.; Koyama, M.; and Ishii, S. (2019). Virtual adversarial training: A regularization method for supervised and semi-supervised learning. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 41:1979–1993.

[221] Mo, S.; Sun, Z.; and Li, C. (2023). Multi-level contrastive learning for self-supervised vision transformers. In *2023 IEEE/CVF Winter Conference on Applications of Computer Vision (WACV)*, pages 2777–2786.

[222] Mopuri, K. R. and Babu, R. V. (2015). Object level deep feature pooling for compact image representation. In *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR) Workshops*.

[223] Moran, N.; Schmidt, D.; Zhong, Y.; and Coady, P. (2020). Noisier2noise: Learning to denoise from unpaired noisy data. In *2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 12061–12069.

[224] Mou, Y.; He, K.; Wang, P.; Wu, Y.; Wang, J.; Wu, W.; and Xu, W. (2022). Watch the neighbors: A unified k-nearest neighbor contrastive learning framework for ood intent discovery.

[225] Nguyen, V.; Ngo, T. D.; Nguyen, K. M. T. T.; Duong, D. A.; Nguyen, K.; and Le, D. (2013). Re-ranking for person re-identification. In *2013 International Conference on Soft Computing and Pattern Recognition (SoCPaR)*, pages 304–308.

[226] Ni, H.; Li, Y.; Gao, L.; Shen, H.; and Song, J. (2023). Part-aware transformer for generalizable person re-identification. In *2023 IEEE/CVF International Conference on Computer Vision (ICCV)*, pages 11246–11255, Los Alamitos, CA, USA. IEEE Computer Society.

[227] Nie, L.; Jiao, F.; Wang, W.; Wang, Y.; and Tian, Q. (2021). Conversational image search. *IEEE Transactions on Image Processing*, 30:7732–7743.

[228] Nie, L.; Li, Y.; Feng, F.; Song, X.; Wang, M.; and Wang, Y. (2020). Large-scale question tagging via joint question-topic embedding learning. *ACM Trans. Inf. Syst.*, 38(2).

[229] Nilsback, M.-E. and Zisserman, A. (2006). A visual vocabulary for flower classification. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, volume 2, pages 1447–1454.

[230] Nistér, D. and Stewénius, H. (2006). Scalable recognition with a vocabulary tree. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR'2006)*, volume 2, pages 2161–2168.

[231] Ojala, T.; Pietikäinen, M.; and Mäenpää, T. (2002). Multiresolution gray-scale and rotation invariant texture classification with local binary patterns. *PAMI*, 24(7):971–987.

[232] Oliva, A. and Torralba, A. (2001). Modeling the shape of the scene: A holistic representation of the spatial envelope. *IJCV*, 42(3):145–175.

[233] Oliveira, A. and Rocha, A. (2020). Contextual features and sequence labeling techniques for relevance prediction in retrieval. In *2020 International Conference on Systems, Signals and Image Processing (IWSSIP)*, pages 305–310.

[234] Omohundro, S. M. (1989). Five balltree construction algorithms. Technical report, International Computer Science Institute.

[235] Ouyang, J.; Wu, H.; Wang, M.; Zhou, W.; and Li, H. (2021). Contextual similarity aggregation with self-attention for visual re-ranking. In Ranzato, M.; Beygelzimer, A.; Dauphin, Y.; Liang, P.; and Vaughan, J. W., editors, *Advances in Neural Information Processing Systems*, volume 34, pages 3135–3148. Curran Associates, Inc.

[236] Papa, J. P.; Falcão, A.; and Suzuki, C. (2009). Supervised pattern classification based on optimum-path forest. *Int. J. Imaging Syst. Technol.*, 19(2):120–131.

[237] Park, G.; Baek, Y.; and Lee, H.-K. (2005). Re-ranking algorithm using post-retrieval clustering for content-based image retrieval. *Information Processing and Management*, 41(2):177–194.

[238] Paulin, M.; Mairal, J.; Douze, M.; Harchaoui, Z.; Perronnin, F.; and Schmid, C. (2017). Convolutional patch representations for image retrieval: An unsupervised approach. *Int. Journal of Computer Vision*, 121(1):149–168.

[239] Pedregosa, F.; Varoquaux, G.; Gramfort, A.; Michel, V.; Thirion, B.; Grisel, O.; Blondel, M.; Prettenhofer, P.; Weiss, R.; Dubourg, V.; Vanderplas, J.; Passos, A.; Cournapeau, D.; Brucher, M.; Perrot, M.; and Duchesnay, E. (2011). Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research*, 12:2825–2830.

[240] Pedronette, D. C. G.; Almeida, J.; and Torres, R. d. S. (2014a). A scalable re-ranking method for content-based image retrieval. *Information Sciences*, 265:91–104.

[241] Pedronette, D. C. G.; Gonçalves, F. M. F.; and Guilherme, I. R. (2018). Unsupervised manifold learning through reciprocal kNN graph and Connected Components for image retrieval tasks. *Pattern Recognition*, 75:161 – 174.

[242] Pedronette, D. C. G. and Latecki, L. J. (2021). Rank-based self-training for graph convolutional networks. *Information Processing & Management*, 58(2):102443.

[243] Pedronette, D. C. G.; Penatti, O. A.; and Torres, R. d. S. (2014b). Unsupervised manifold learning using reciprocal knn graphs in image re-ranking and rank aggregation tasks. *Image and Vision Computing*, 32(2):120 – 130.

[244] Pedronette, D. C. G. and Torres, R. d. S. (2010). Shape retrieval using contour features and distance optmization. In *VISAPP*, volume 1, pages 197 – 202.

[245] Pedronette, D. C. G. and Torres, R. d. S. (2011). Exploiting contextual spaces for image re-ranking and rank aggregation. In *Proceedings of the 1st International Conference on Multimedia Retrieval, (ICMR)*, page 13. ACM.

[246] Pedronette, D. C. G. and Torres, R. d. S. (2012). Exploiting contextual information for image re-ranking and rank aggregation. *International Journal of Multimedia Information Retrieval*, 1(2):115–128.

[247] Pedronette, D. C. G. and Torres, R. d. S. (2013). Image re-ranking and rank aggregation based on similarity of ranked lists. *Pattern Recognition*, 46(8):2350–2360.

[248] Pedronette, D. C. G. and Torres, R. d. S. (2015). Unsupervised effectiveness estimation for image retrieval using reciprocal rank information. In *2015 28th SIBGRAPI Conference on Graphics, Patterns and Images*, pages 321–328.

[249] Pedronette, D. C. G. and Torres, R. d. S. (2016). A correlation graph approach for unsupervised manifold learning in image retrieval tasks. *Neurocomputing*, 208:66–79. SI: BridgingSemantic.

[250] Pedronette, D. C. G.; Torres, R. d. S.; and Calumby, R. T. (2014c). Using contextual spaces for image re-ranking and rank aggregation. *Multimedia Tools and Applications*, 69(3):689–716.

[251] Pedronette, D. C. G.; Valem, L. P.; Almeida, J.; and Torres, R. d. S. (2019). Multimedia retrieval through unsupervised hypergraph-based manifold ranking. *IEEE Transactions on Image Processing*, 28(12):5824–5838.

[252] Pedronette, D. C. G.; Valem, L. P.; and Latecki, L. J. (2021a). Efficient rank-based diffusion process with assured convergence. *Journal of Imaging*, 7(3).

[253] Pedronette, D. C. G.; Valem, L. P.; and Torres, R. d. S. (2021b). A bfs-tree of ranking references for unsupervised manifold learning. *Pattern Recognition*, 111:107666.

[254] Penatti, O. A.; Valle, E.; and Torres, R. d. S. (2012). Comparative study of global color and texture descriptors for web image retrieval. *Journal of Visual Communication and Image Representation*, 23(2):359–380.

[255] Peng, L.; Qiang, B.; and Wu, J. (2022). A survey: Image classification models based on convolutional neural networks. In *2022 14th International Conference on Computer Research and Development (ICCRD)*, pages 291–298.

[256] Pereira-Ferrero, V.; Lewis, T.; Valem, L.; Ferrero, L.; Pedronette, D.; and Latecki, L. (2024a). Unsupervised affinity learning based on manifold analysis for image retrieval: A survey. *Computer Science Review*, 53:100657.

[257] Pereira-Ferrero, V.; Valem, L.; Leticio, G.; and Pedronette, D. (2023a). Feature fusion and augmentation based on manifold ranking for image classification. In *2023 IEEE Sixth International Conference on Artificial Intelligence and Knowledge Engineering (AIKE)*, pages 75–82, Los Alamitos, CA, USA. IEEE Computer Society.

[258] Pereira-Ferrero, V. H.; Valem, L. P.; Leticio, G. R.; and Pedronette, D. C. G. (2024b). Feature fusion and augmentation based on manifold ranking for image classification. *International Journal of Semantic Computing (IJSC)*. Accepted, to appear.

[259] Pereira-Ferrero, V. H.; Valem, L. P.; and Pedronette, D. C. G. (2023b). Feature augmentation based on manifold ranking and lstm for image classification. *Expert Systems with Applications*, 213:118995.

[260] Piras, L. and Giacinto, G. (2017). Information fusion in content based image retrieval: A comprehensive overview. *Information Fusion*, 37:50–60.

[261] Pisani, F.; Pascotti Valem, L.; Guimarães Pedronette, D. C.; da S. Torres, R.; Borin, E.; and Breternitz Jr., M. (2020). A unified model for accelerating unsupervised iterative

re-ranking algorithms. *Concurrency and Computation: Practice and Experience*, 32(14):e5702.

[262] Poesina, E.; Ionescu, R. T.; and Mothe, J. (2023). iQPP: A benchmark for image query performance prediction. In *Proceedings of the 46th International ACM SIGIR Conference on Research and Development in Information Retrieval*, SIGIR '23, page 2953–2963, New York, NY, USA. Association for Computing Machinery.

[263] Presotto, J. G. C.; Dos Santos, S. F.; Valem, L. P.; Faria, F. A.; Papa, J. P.; Almeida, J.; and Pedronette, D. C. G. (2022). Weakly supervised learning based on hypergraph manifold ranking. *Journal of Visual Communication and Image Representation*, 89:103666.

[264] Presotto, J. G. C.; Valem, L. P.; De Sá, N. G.; Pedronette, D. C. G.; and Papa, J. P. (2021). Weakly supervised learning through rank-based contextual measures. In *2020 25th International Conference on Pattern Recognition (ICPR)*, pages 5752–5759.

[265] Presotto, J. G. C.; Valem, L. P.; De Sá, N. G.; Pedronette, D. C. G.; and Papa, J. P. (2024). Weakly supervised classification through manifold learning and rank-based contextual measures. *Neurocomputing*, 589(C).

[266] Presotto, J. G. C.; Valem, L. P.; and Pedronette, D. C. G. (2019). Unsupervised effectiveness estimation through intersection of ranking references. In *Computer Analysis of Images and Patterns - CAIP 2019*, volume 11679, pages 231–244.

[267] Qin, D.; Gammeter, S.; Bossard, L.; Quack, T.; and van Gool, L. (2011). Hello neighbor: Accurate object retrieval with k-reciprocal nearest neighbors. In *CVPR 2011*, pages 777–784.

[268] Qin, D.; Wengert, C.; and Gool, L. V. (2013). Query adaptive similarity for large scale object retrieval. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR'2013)*, pages 1610–1617.

[269] Quan, Y.; Chen, M.; Pang, T.; and Ji, H. (2020). Self2self with dropout: Learning self-supervised denoising from single image. In *2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 1887–1895.

[270] Raghu, M. and Schmidt, E. (2020). A survey of deep learning for scientific discovery. *arXiv preprint arXiv:2003.11755*.

[271] Razavian, A. S.; Sullivan, J.; Carlsson, S.; and Maki, A. (2016). Visual instance retrieval with deep convolutional networks. *ITE Transactions on Media Technology and Applications*, 4(3):251–258.

[272] Redmon, J.; Divvala, S.; Girshick, R.; and Farhadi, A. (2016). You only look once: Unified, real-time object detection. In *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 779–788.

[273] Ren, C.; Liang, B.; and Lei, Z. (2019). Domain adaptive person re-identification via camera style generation and label propagation. *CoRR*, abs/1905.05382.

[274] Rodrigues, J. and Carbonera, J. (2024). Graph convolutional networks for image classification: Comparing approaches for building graphs from images. In *Proceedings of the 26th International Conference on Enterprise Information Systems - Volume 1: ICEIS*, pages 437–446. INSTICC, SciTePress.

[275] Sandler, M.; Howard, A.; Zhu, M.; Zhmoginov, A.; and Chen, L. (2018). Mobilenetv2: Inverted residuals and linear bottlenecks. In *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 4510–4520.

[276] Schmid, C. (2001). Constructing models for content-based image retrieval. In *Proceedings of the 2001 IEEE Computer Society Conference on Computer Vision and Pattern Recognition. CVPR 2001*, volume 2, pages II–II.

[277] Schölkopf, B.; Platt, J.; and Hofmann, T. (2007). Learning with hypergraphs: Clustering, classification, and embedding. In *Advances in Neural Information Processing Systems (NIPS'07)*, pages 1601–1608.

[278] Shao, P. and Tao, J. (2024). Multi-level graph contrastive learning. *Neurocomputing*, 570:127101.

[279] Shao, S.; Chen, K.; Karpur, A.; Cui, Q.; Araujo, A.; and Cao, B. (2023). Global features are all you need for image retrieval and reranking. In *2023 IEEE/CVF International Conference on Computer Vision (ICCV)*, pages 11002–11012.

[280] Sharma, A. and Kumar, D. (2022). Classification with 2-d convolutional neural networks for breast cancer diagnosis. *Scientific Reports*, 12(1):21857.

[281] Sharma, A.; Vans, E.; Shigemizu, D.; Boroevich, K. A.; and Tsunoda, T. (2019). Deepinsight: A methodology to transform a non-image data to an image for convolution neural network architecture. *Scientific Reports*, 9(1):11399.

[282] Shen, X.; Xiao, Y.; Hu, S. X.; Sbai, O.; and Aubry, M. (2021). Re-ranking for image retrieval and transductive few-shot classification. In Ranzato, M.; Beygelzimer, A.; Dauphin, Y.; Liang, P.; and Vaughan, J. W., editors, *Advances in Neural Information Processing Systems*, volume 34, pages 25932–25943.

[283] Shimomura, L.; Oyamada, R.; Vieira, M.; and Kaster, D. (2021). A survey on graph-based methods for similarity searches in metric spaces. *Information Systems*, 95:101507.

[284] Shriram K. Vasudevan, P.L.K. Priyadarsini, S. V. (2012). *Content Based Image Retrieval (CBIR): A deeper insight*. LAP LAMBERT Academic Publishing.

[285] Shtok, A.; Kurland, O.; and Carmel, D. (2016). Query performance prediction using reference lists. *ACM Trans. Inf. Syst.*, 34(4).

[286] Shu, Y.; Gu, X.; Yang, G.-Z.; and Lo, B. P. L. (2022). Revisiting self-supervised contrastive learning for facial expression recognition. In *33rd British Machine Vision Conference 2022, BMVC 2022, London, UK, November 21-24, 2022*. BMVA Press.

[287] Siam, M. K. H.; Bhattacharjee, M.; Mahmud, S.; Sarkar, M. S.; and Rana, M. M. (2024). The impact of machine learning on society: An analysis of current trends and future implications. *arXiv preprint arXiv:2404.10204*.

[288] Singh, N. K.; Khare, M.; and Jethva, H. B. (2022). A comprehensive survey on person re-identification approaches: various aspects. *Multimedia Tools and Applications*, 81(11):15747–15791.

[289] Sohn, K.; Berthelot, D.; Li, C.-L.; Zhang, Z.; Carlini, N.; Cubuk, E. D.; Kurakin, A.; Zhang, H.; and Raffel, C. (2020). Fixmatch: Simplifying semi-supervised learning with consistency and confidence. In *Proceedings of the 34th International Conference on Neural Information Processing Systems*, NIPS'20, Red Hook, NY, USA. Curran Associates Inc.

[290] Soltanayev, S. and Chun, S. Y. (2018). Training deep learning based denoisers without ground truth data. In Bengio, S.; Wallach, H.; Larochelle, H.; Grauman, K.; Cesa-Bianchi, N.; and Garnett, R., editors, *Advances in Neural Information Processing Systems 31*, pages 3257–3267. Curran Associates, Inc.

[291] Somers, V.; Vleeschouwer, C. D.; and Alahi, A. (2023). Body part-based representation learning for occluded person re-identification. In *2023 IEEE/CVF Winter Conference on Applications of Computer Vision (WACV)*, pages 1613–1623.

[292] Song, K.; Han, J.; Cheng, G.; Lu, J.; and Nie, F. (2022). Adaptive neighborhood metric learning. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 44(9):4591–4604.

[293] Song, Z.; Yang, X.; Xu, Z.; and King, I. (2023). Graph-based semi-supervised learning: A comprehensive review. *IEEE Transactions on Neural Networks and Learning Systems*, 34(11):8174–8194.

[294] Srivastava, D.; Singh, S. S.; Rajitha, B.; Verma, M.; Kaur, M.; and Lee, H.-N. (2023). Content-based image retrieval: A survey on local and global features selection, extraction, representation, and evaluation parameters. *IEEE Access*, 11:95410–95431.

[295] Stehling, R. O.; Nascimento, M. A.; and Falcão, A. X. (2002). A compact and efficient image retrieval approach based on border/interior pixel classification. In *CIKM*, pages 102–109.

[296] Sukhbaatar, S.; Bruna, J.; Paluri, M.; Bourdev, L.; and Fergus, R. (2014). Training convolutional networks with noisy labels. *arXiv preprint arXiv:1406.2080*.

[297] Sun, H.; Li, M.; and Li, C.-G. (2022). Hybrid contrastive learning with cluster ensemble for unsupervised person re-identification. In Wallraven, C.; Liu, Q.; and Nagahara, H., editors, *Pattern Recognition*, pages 532–546, Cham. Springer International Publishing.

[298] Sun, L.; Ji, S.; and Ye, J. (2008). Hypergraph spectral learning for multi-label classification. In *ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, KDD '08, pages 668–676.

[299] Sun, S.; Li, Y.; Zhou, W.; Tian, Q.; and Li, H. (2017). Local residual similarity for image re-ranking. *Information Sciences*, 417(Sup. C):143 – 153.

[300] Sun, S.; Zhou, W.; Tian, Q.; Yang, M.; and Li, H. (2018a). Assessing image retrieval quality at the first glance. *IEEE Transactions on Image Processing*, 27(12):6124–6134.

[301] Sun, Y.; Zheng, L.; Yang, Y.; Tian, Q.; and Wang, S. (2018b). Beyond part models: Person retrieval with refined part pooling (and a strong convolutional baseline). In Ferrari, V.; Hebert, M.; Sminchisescu, C.; and Weiss, Y., editors, *Computer Vision – ECCV 2018*, pages 501–518, Cham. Springer International Publishing.

[302] Szegedy, C.; Ioffe, S.; Vanhoucke, V.; and Alemi, A. A. (2017). Inception-v4, inception-resnet and the impact of residual connections on learning. In *AAAI*, volume 4, page 12.

[303] Sánchez, J.; Perronnin, F.; Mensink, T.; and Verbeek, J. (2013). Image classification with the fisher vector: Theory and practice. *International Journal of Computer Vision*, 105.

[304] Taigman, Y.; Yang, M.; Ranzato, M.; and Wolf, L. (2014). Deepface: Closing the gap to human-level performance in face verification. In *2014 IEEE Conference on Computer Vision and Pattern Recognition*, pages 1701–1708.

[305] Tan, F.; Yuan, J.; and Ordonez, V. (2021). Instance-level image retrieval using reranking transformers. In *2021 IEEE/CVF International Conference on Computer Vision (ICCV)*, pages 12085–12095.

[306] Tang, H.; Zhao, Y.; and Lu, H. (2019). Unsupervised person re-identification with iterative self-supervised domain adaptation. In *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR) Workshops*.

[307] Tang, W. (2023). Review of image classification algorithms based on graph convolutional networks. *EAI Endorsed Transactions on AI and Robotics*, 2(1).

[308] Tao, B. and Dickinson, B. W. (2000). Texture recognition and image retrieval using gradient indexing. *JVCIR*, 11(3):327–342.

[309] Tarvainen, A. and Valpola, H. (2017). Mean teachers are better role models: Weight-averaged consistency targets improve semi-supervised deep learning results. In *Proceedings of the 31st International Conference on Neural Information Processing Systems*, NIPS'17, page 1195–1204, Red Hook, NY, USA. Curran Associates Inc.

[310] Tay, C.; Roy, S.; and Yap, K. (2019). Aanet: Attribute attention network for person re-identifications. In *2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 7127–7136.

[311] Thammasorn, P.; Hippe, D. S.; Chaovalitwongse, W. A.; Spraker, M.; Wootton, L.; Nyflot, M.; Combs, S.; Peeken, J.; and Ford, E. (2019). Neighborhood watch: Representation learning with local-margin triplet loss and sampling strategy for k-nearest-neighbor image classification. *CoRR*, abs/1911.07940.

[312] Thomee, B. and Lew, M. S. (2012). Interactive search in image retrieval: a survey. *International Journal of Multimedia Information Retrieval*, 1(2):71–86.

[313] Tian, X.; Jia, Q.; and Mei, T. (2015). Query difficulty estimation for image search with query reconstruction error. *IEEE Transactions on Multimedia*, 17(1):79–91.

[314] Tian, X.; Lu, Y.; and Yang, L. (2012). Query difficulty prediction for web image search. *IEEE Transactions on Multimedia*, 14(4):951–962.

[315] Tolias, G.; Avrithis, Y.; and Jégou, H. (2013). To aggregate or not to aggregate: Selective match kernels for image search. In *IEEE International Conference on Computer Vision (ICCV'2013)*, pages 1401–1408.

[316] Torres, R. d. S. and Falcão, A. X. (2006). Content-based image retrieval: Theory and applications. *Revista de Informática Teórica e Aplicada*, 13:161–185.

[317] Torres, R. d. S. and Falcão, A. X. (2007). Contour Salience Descriptors for Effective Image Retrieval and Analysis. *Image and Vision Computing*, 25(1):3–13.

[318] Torres, R. d. S.; Falcão, A. X.; Gonçalves, M. A.; Papa, J. P.; Zhang, B.; Fan, W.; and Fox, E. A. (2009). A genetic programming framework for content-based image

retrieval. *Pattern Recognition*, 42(2):283–292. Learning Semantics from Multimedia Content.

[319] Toshev, A. and Szegedy, C. (2013). Deeppose: Human pose estimation via deep neural networks. *CoRR*, abs/1312.4659.

[320] Tripathi, S. and King, C. R. (2024). Contrastive learning: Big data foundations and applications. In *Proceedings of the 7th Joint International Conference on Data Science & Management of Data (11th ACM IKDD CODS and 29th COMAD)*, CODS-COMAD '24, page 493–497, New York, NY, USA. Association for Computing Machinery.

[321] Uelwer, T.; Robine, J.; Wagner, S. S.; Höftmann, M.; Upschulte, E.; Konietzny, S.; Behrendt, M.; and Harmeling, S. (2023). A survey on self-supervised representation learning. *arXiv preprint arXiv:2308.11455*.

[322] Urbanowicz, R. J.; Meeker, M.; La Cava, W.; Olson, R. S.; and Moore, J. H. (2018). Relief-based feature selection: Introduction and review. *Journal of Biomedical Informatics*, 85:189–203.

[323] Valem, L. P.; De Oliveira, C. R.; Pedronette, D. C. G.; and Almeida, J. (2018a). Unsupervised similarity learning through rank correlation and knn sets. *ACM Trans. Multim. Comput. Commun. Appl.*, 14(4):80:1–80:23.

[324] Valem, L. P.; Kawai, V. A. S.; Pereira-Ferrero, V. H.; and Pedronette, D. C. G. (2022). A novel rank correlation measure for manifold learning on image retrieval and person re-id. In *2022 IEEE International Conference on Image Processing (ICIP)*, pages 1371–1375.

[325] Valem, L. P.; Pedronette, D.; and Allili, M. S. (2024). Contrastive loss based on contextual similarity for image classification. In *19th International Symposium on Visual Computing (ISVC)*. Submitted.

[326] Valem, L. P. and Pedronette, D. C. G. (2017). An unsupervised distance learning framework for multimedia retrieval. In *Proceedings of the 2017 ACM on International Conference on Multimedia Retrieval*, ICMR '17, pages 107–111, New York, NY, USA. ACM.

[327] Valem, L. P. and Pedronette, D. C. G. (2019). An unsupervised genetic algorithm framework for rank selection and fusion on image retrieval. In *Proceedings of the 2019 on International Conference on Multimedia Retrieval*, ICMR '19, page 58–62, New York, NY, USA. Association for Computing Machinery.

[328] Valem, L. P. and Pedronette, D. C. G. (2020a). Graph-based selective rank fusion for unsupervised image retrieval. *Pattern Recognition Letters*, 135:82–89.

[329] Valem, L. P. and Pedronette, D. C. G. (2020b). Unsupervised selective rank fusion for image retrieval tasks. *Neurocomputing*, 377:182 – 199.

[330] Valem, L. P. and Pedronette, D. C. G. (2021). A denoising convolutional neural network for self-supervised rank effectiveness estimation on image retrieval. In *Proceedings of the 2021 International Conference on Multimedia Retrieval*, ICMR '21, page 294–302, New York, NY, USA. Association for Computing Machinery.

[331] Valem, L. P. and Pedronette, D. C. G. (2022). Person re-id through unsupervised hypergraph rank selection and fusion. *Image Vision Comput.*, 123(C).

[332] Valem, L. P.; Pedronette, D. C. G.; and Almeida, J. (2018b). Unsupervised similarity learning through cartesian product of ranking references. *Pattern Recognition Letters*, 114:41 – 52.

[333] Valem, L. P.; Pedronette, D. C. G.; and Latecki, L. J. (2023a). Graph convolutional networks based on manifold learning for semi-supervised image classification. *Computer Vision and Image Understanding*, 227:103618.

[334] Valem, L. P.; Pedronette, D. C. G.; and Latecki, L. J. (2023b). Rank flow embedding for unsupervised and semi-supervised manifold learning. *IEEE Transactions on Image Processing*, 32:2811–2826.

[335] Valem, L. P.; Pedronette, D. C. G.; Torres, R. d. S.; Borin, E.; and Almeida, J. (2015). Effective, efficient, and scalable unsupervised distance learning in image retrieval tasks. In *Proceedings of the 5th ACM on International Conference on Multimedia Retrieval*, ICMR '15, pages 51–58, New York, NY, USA. ACM.

[336] Valem, L. P.; Pereira-Ferrero, V.; and Pedronette, D. (2023c). Self-supervised regression for query performance prediction on image retrieval. In *2023 IEEE Sixth International Conference on Artificial Intelligence and Knowledge Engineering (AIKE)*, pages 95–98, Los Alamitos, CA, USA. IEEE Computer Society.

[337] Van de Sande, K. E. A.; Gevers, T.; and Snoek, C. G. M. (2010). Evaluating color descriptors for object and scene recognition. *PAMI*, 32(9):1582–1596.

[338] Vanyan, A. and Khachatrian, H. (2021). Deep semi-supervised image classification algorithms: a survey. *JUCS - Journal of Universal Computer Science*, 27(12):1390–1407.

[339] Varior, R. R.; Haloi, M.; and Wang, G. (2016a). Gated siamese convolutional neural network architecture for human re-identification. *ArXiv*, abs/1607.08378.

[340] Varior, R. R.; Shuai, B.; Lu, J.; Xu, D.; and Wang, G. (2016b). A siamese long short-term memory architecture for human re-identification. In *ECCV*.

[341] Varoquaux, G. and Cheplygina, V. (2022). Machine learning for medical imaging: methodological failures and recommendations for the future. *npj Digital Medicine*, 5(1):48.

[342] Vassou, S. A.; Anagnostopoulos, N.; Amanatiadis, A.; Christodoulou, K.; and Chatzichristofis, S. A. (2017). Como: A compact composite moment-based descriptor for image retrieval. In *Proceedings of the 15th International Workshop on Content-Based Multimedia Indexing*, CBMI '17, pages 30:1–30:5, New York, NY, USA. ACM.

[343] Vasudevan, V.; Bassenne, M.; Islam, M. T.; and Xing, L. (2023). Image classification using graph neural network and multiscale wavelet superpixels. *Pattern Recogn. Lett.*, 166(C):89–96.

[344] Veličković, P.; Cucurull, G.; Casanova, A.; Romero, A.; Liò, P.; and Bengio, Y. (2018). Graph attention networks. In *International Conference on Learning Representations*.

[345] Vinyals, O.; Blundell, C.; Lillicrap, T.; Wierstra, D.; and Others (2016). Matching networks for one shot learning. In *Advances in Neural Information Processing Systems*, pages 3630–3638.

[346] Wah, C.; Branson, S.; Welinder, P.; Perona, P.; and Belongie, S. (2011). The Caltech-UCSD Birds-200-2011 Dataset. Technical Report CNS-TR-2011-001, California Institute of Technology.

[347] Wang, B. and Tu, Z. (2012). Affinity learning via self-diffusion for image segmentation and clustering. In *2012 IEEE CVPR*, pages 2312–2319.

[348] Wang, D. and Zhang, S. (2020). Unsupervised person re-identification via multi-label classification. In *CVPR*, pages 10978–10987.

[349] Wang, G.; Wang, G.; Zhang, X.; Lai, J.; Yu, Z.; and Lin, L. (2020). Weakly supervised person re-id: Differentiable graphical learning and a new benchmark.

[350] Wang, H.; Zhu, X.; Xiang, T.; and Gong, S. (2016). Towards unsupervised open-set person re-identification. In *2016 IEEE International Conference on Image Processing (ICIP)*, pages 769–773.

[351] Wang, J.; Li, Y.; Bai, X.; Zhang, Y.; Wang, C.; and Tang, N. (2011a). Learning context-sensitive similarity by shortest path propagation. *Pattern Recognition*, 44(10-11):2367–2374.

[352] Wang, J. and Zhu, Z. (2010). Image retrieval system based on multi-feature fusion and relevance feedback. In *2010 International Conference on Machine Learning and Cybernetics*, volume 4, pages 2053–2058.

[353] Wang, M.; Lai, B.; Huang, J.; Gong, X.; and Hua, X.-S. (2021). Camera-aware proxies for unsupervised person re-identification. In *Proceedings of the AAAI Conference on Artificial Intelligence (AAAI)*.

[354] Wang, M.; Li, J.; Lai, B.; Gong, X.; and Hua, X.-S. (2022). Offline-online associated camera-aware proxies for unsupervised person re-identification. *IEEE Transactions on Image Processing*, 31:6548–6561.

[355] Wang, M. and Song, T. (2013). Remote sensing image retrieval by scene semantic matching. *IEEE Transactions on Geoscience and Remote Sensing*, 51(5):2874–2886.

[356] Wang, X.; Yang, M.; Cour, T.; Zhu, S.; Yu, K.; and Han, T. (2011b). Contextual weighting for vocabulary tree based image retrieval. In *ICCV'2011*, pages 209–216.

[357] Webber, C.; Ioana, C.; and Marwan, N. (2016). *Recurrence Plots and Their Quantifications: Expanding Horizons: Proceedings of the 6th International Symposium on Recurrence Plots, Grenoble, France, 17-19 June 2015*. Springer.

[358] Webber, W.; Moffat, A.; and Zobel, J. (2010). A similarity measure for indefinite rankings. *ACM Transactions on Information Systems*, 28(4):20:1–20:38.

[359] Wei, L.; Zhang, S.; Gao, W.; and Tian, Q. (2018). Person transfer gan to bridge domain gap for person re-identification. In *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 79–88.

[360] Weinberger, K. Q.; Blitzer, J.; and Saul, L. (2006). Distance metric learning for large margin nearest neighbor classification. In Weiss, Y.; Schölkopf, B.; and Platt, J., editors, *Advances in Neural Information Processing Systems*, volume 18, pages 1473–1480. MIT Press.

[361] Wojke, N.; Bewley, A.; and Paulus, D. (2017). Simple online and realtime tracking with a deep association metric. In *2017 IEEE International Conference on Image Processing (ICIP)*, pages 3645–3649. IEEE.

[362] Wu, C.; Wang, C.; Xu, J.; Liu, Z.; Zheng, K.; Wang, X.; Song, Y.; and Gai, K. (2023a). Graph contrastive learning with generative adversarial network. In *Proceedings of the 29th ACM SIGKDD Conference on Knowledge Discovery and Data Mining*, KDD '23, page 2721–2730, New York, NY, USA. Association for Computing Machinery.

[363] Wu, F.; Souza, A.; Zhang, T.; Fifty, C.; Yu, T.; and Weinberger, K. (2019). Simplifying graph convolutional networks. In Chaudhuri, K. and Salakhutdinov, R., editors, *Proceedings of the 36th International Conference on Machine Learning*, volume 97 of *Proceedings of Machine Learning Research*, pages 6861–6871. PMLR.

[364] Wu, H.; Wang, M.; Zhou, W.; Lu, Z.; and Li, H. (2023b). Asymmetric feature fusion for image retrieval. In *2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 11082–11092.

[365] Wu, L.; Shen, C.; and van den Hengel, A. (2016). Deep linear discriminant analysis on fisher networks: A hybrid architecture for person re-identification. *CoRR*, abs/1606.01595.

[366] Wu, Y.; Bourahla, O. E. F.; Li, X.; Wu, F.; Tian, Q.; and Zhou, X. (2020). Adaptive graph representation learning for video person re-identification. *IEEE Transactions on Image Processing*, 29:8821–8830.

[367] Wu, Y.; Huang, T.; Yao, H.; Zhang, C.; Shao, Y.; Han, C.; Gao, C.; and Sang, N. (2022). Multi-centroid representation network for domain adaptive person re-id. *Proceedings of the AAAI Conference on Artificial Intelligence*, 36(3):2750–2758.

[368] Wu, Y.; Wu, X.; Li, X.; and Tian, J. (2021). Mgh: Metadata guided hypergraph modeling for unsupervised person re-identification. In *Proceedings of the 29th ACM International Conference on Multimedia*, MM '21, page 1571–1580, New York, NY, USA. Association for Computing Machinery.

[369] Xia, F.; Sun, K.; Yu, S.; Aziz, A.; Wan, L.; Pan, S.; and Liu, H. (2021). Graph learning: A survey. *IEEE Transactions on Artificial Intelligence*, 2(02):109–127.

[370] Xian, Y. and Hu, H. (2018). Enhanced multi-dataset transfer learning method for unsupervised person re-identification using co-training strategy. *IET Computer Vision*, 12(8):1219–1227.

[371] Xie, L.; Hong, R.; Zhang, B.; and Tian, Q. (2015). Image classification and retrieval are one. In *ACM ICMR'2015*, pages 3–10.

[372] Xie, S.; Girshick, R.; Dollár, P.; Tu, Z.; and He, K. (2017). Aggregated residual transformations for deep neural networks. In *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 5987–5995.

[373] Xin, X.; Wang, J.; Xie, R.; Zhou, S.; Huang, W.; and Zheng, N. (2019). Semi-supervised person re-identification using multi-view clustering. *Pattern Recognition*, 88:285–297.

[374] Xing, X.; Zhang, Y.; and Han, M. (2010). Query difficulty prediction for contextual image retrieval. In *European Conference on IR Research, ECIR*, volume 5993, pages 581–585.

[375] Xiong, F.; Gou, M.; Camps, O.; and Sznaier, M. (2014). Person re-identification using kernel-based metric learning methods. In Fleet, D.; Pajdla, T.; Schiele, B.; and

Tuytelaars, T., editors, *Computer Vision – ECCV 2014*, pages 1–16, Cham. Springer International Publishing.

[376] Xuan, S. and Zhang, S. (2021). Intra-inter camera similarity for unsupervised person re-identification. In *2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 11921–11930.

[377] Yang, B.; Shan, Y.; Peng, R.; Li, J.; Chen, S.; and Li, L. (2022). A feature extraction method for person re-identification based on a two-branch cnn. *Multimedia Tools and Applications*, 81(27):39169–39184.

[378] Yang, F.; Hinami, R.; Matsui, Y.; Ly, S.; and Satoh, S. (2019a). Efficient image retrieval via decoupling diffusion into online and offline processing. In *Conference on Artificial Intelligence, AAAI 2019*, pages 9087–9094. AAAI Press.

[379] Yang, J.; Zheng, W.-S.; Yang, Q.; Chen, Y.-C.; and Tian, Q. (2020). Spatial-temporal graph convolutional network for video-based person re-identification. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 3286–3296.

[380] Yang, Q.; Yu, H.-X.; Wu, A.; and Zheng, W.-S. (2019b). Patch-based discriminative feature learning for unsupervised person re-identification. In *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*.

[381] Yang, X.; Bai, X.; Latecki, L. J.; and Tu, Z. (2008). Improving shape retrieval by learning graph transduction. In *ECCV*, volume 4, pages 788–801.

[382] Yang, X.; Koknar-Tezel, S.; and Latecki, L. J. (2009a). Locally constrained diffusion process on locally densified distance spaces with applications to shape retrieval. In *CVPR*, pages 357–364.

[383] Yang, X.; Koknar-Tezel, S.; and Latecki, L. J. (2009b). Locally constrained diffusion process on locally densified distance spaces with applications to shape retrieval. In *2009 IEEE Conference on Computer Vision and Pattern Recognition*, pages 357–364.

[384] Yang, X. and Latecki, L. J. (2011). Affinity learning on a tensor product graph with applications to shape and image retrieval. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR'2011)*, pages 2369–2376.

[385] Yang, X.; Prasad, L.; and Latecki, L. J. (2013). Affinity learning with diffusion on tensor product graph. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 35(1):28–38.

[386] Yang, X.; Zhou, P.; and Wang, M. (2019). Person reidentification via structural deep metric learning. *IEEE Transactions on Neural Networks and Learning Systems*, 30(10):2987–2998.

[387] Yang, Y.; Yang, J.; Yan, J.; Liao, S.; Yi, D.; and Li, S. Z. (2014). Salient color names for person re-identification. In Fleet, D.; Pajdla, T.; Schiele, B.; and Tuytelaars, T., editors, *Computer Vision – ECCV 2014*, pages 536–551, Cham. Springer International Publishing.

[388] Ye, M.; Chen, J.; Leng, Q.; Liang, C.; Wang, Z.; and Sun, K. (2015). Coupled-view based ranking optimization for person re-identification. In He, X.; Luo, S.; Tao, D.; Xu, C.; Yang, J.; and Hasan, M. A., editors, *MultiMedia Modeling*, pages 105–117, Cham. Springer International Publishing.

[389] Ye, M.; Liang, C.; Yu, Y.; Wang, Z.; Leng, Q.; Xiao, C.; Chen, J.; and Hu, R. (2016). Person reidentification via ranking aggregation of similarity pulling and dissimilarity pushing. *IEEE Transactions on Multimedia*, 18(12):2553–2566.

[390] Ye, M.; Shen, J.; Lin, G.; Xiang, T.; Shao, L.; and Hoi, S. C. H. (2020). Deep learning for person re-identification: A survey and outlook.

[391] Yi, D.; Lei, Z.; Liao, S.; and Li, S. Z. (2014). Deep metric learning for person re-identification. In *2014 22nd International Conference on Pattern Recognition*, pages 34–39.

[392] You, Y.; Chen, T.; Sui, Y.; Chen, T.; Wang, Z.; and Shen, Y. (2020). Graph contrastive learning with augmentations. In *Proceedings of the 34th International Conference on Neural Information Processing Systems*, NIPS '20, Red Hook, NY, USA. Curran Associates Inc.

[393] Younesi, A.; Ansari, M.; Fazli, M.; Ejlali, A.; Shafique, M.; and Henkel, J. (2024). A comprehensive survey of convolutions in deep learning: Applications, challenges, and future trends. *arXiv preprint arXiv:2402.15490*.

[394] Young, H. (1974). An axiomatization of borda's rule. *Journal of Economic Theory*, 9(1):43–52.

[395] Yu, H.; Wu, A.; and Zheng, W. (2020). Unsupervised person re-identification by deep asymmetric metric embedding. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 42(4):956–973.

[396] Yu, H.-X.; Wu, A.; and Zheng, W.-S. (2017a). Cross-view asymmetric metric learning for unsupervised person re-identification. In *The IEEE International Conference on Computer Vision (ICCV)*.

[397] Yu, H.-X.; Zheng, W.-S.; Wu, A.; Guo, X.; Gong, S.; and Lai, J.-H. (2019). Unsupervised person re-identification by soft multilabel learning. In *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*.

[398] Yu, W.; Yang, K.; Yao, H.; Sun, X.; and Xu, P. (2017b). Exploiting the complementary strengths of multi-layer cnn features for image retrieval. *Neurocomputing*, 237:235–241.

[399] Yuan, L.; Chen, Y.; Wang, T.; Yu, W.; Shi, Y.; Tay, F. E.; Feng, J.; and Yan, S. (2021). Tokens-to-token vit: Training vision transformers from scratch on imagenet. *arXiv preprint arXiv:2101.11986*.

[400] Yuan, Y.; Chen, W.; Yang, Y.; and Wang, Z. (2019). In defense of the triplet loss again: Learning robust person re-identification with fast approximated triplet loss and label distillation. *2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, pages 1454–1463.

[401] Zagoris, K.; Chatzichristofis, S.; Papamarkos, N.; and Boutalis, Y. (2010). Automatic image annotation and retrieval using the joint composite descriptor. In *14th Panhellenic Conference on Informatics (PCI)*, pages 143–147.

[402] Zeng, K.; Ning, M.; Wang, Y.; and Guo, Y. (2020). Hierarchical clustering with hard-batch triplet loss for person re-identification. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*.

[403] Zeng, Q. and Yu, H. (2019). Group affinity guided deep hypergraph model for person re-identification. *Electronics Letters*, 55(4):186–188.

[404] Zhai, Y.; Lu, S.; Ye, Q.; Shan, X.; Chen, J.; Ji, R.; and Tian, Y. (2020). Ad-cluster: Augmented discriminative clustering for domain adaptive person re-identification. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*.

[405] Zhang, G.; Pan, J.; Zhang, Z.; Zhang, H.; Xing, C.; Sun, B.; and Li, M. (2021a). Hybrid graph convolutional network for semi-supervised retinal image classification. *IEEE Access*, 9:35778–35789.

[406] Zhang, H.; Chen, X.; Jing, H.; Zheng, Y.; Wu, Y.; and Jin, C. (2023). Etr: An efficient transformer for re-ranking in visual place recognition. In *2023 IEEE/CVF Winter Conference on Applications of Computer Vision (WACV)*, pages 5654–5663.

[407] Zhang, H.; Li, Y.; Zhuang, Z.; Xie, L.; and Tian, Q. (2021b). 3d-gat: 3d-guided adversarial transform network for person re-identification in unseen domains. *Pattern Recognition*, 112:107799.

[408] Zhang, K.; Li, Y.; Zuo, W.; Zhang, L.; Van Gool, L.; and Timofte, R. (2020). Plug-and-play image restoration with deep denoiser prior. *arXiv preprint*.

[409] Zhang, K.; Zuo, W.; Chen, Y.; Meng, D.; and Zhang, L. (2017). Beyond a gaussian denoiser: Residual learning of deep cnn for image denoising. *IEEE Transactions on Image Processing*, 26(7):3142–3155.

[410] Zhang, L.; Song, H.; Aletras, N.; and Lu, H. (2018a). Graph node-feature convolution for representation learning. *arXiv preprint arXiv:1812.00086*.

[411] Zhang, L.; Xiang, T.; and Gong, S. (2016). Learning a discriminative null space for person re-identification. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*.

[412] Zhang, S.; Tong, H.; Xu, J.; and Maciejewski, R. (2019). Graph convolutional networks: a comprehensive review. *Computational Social Networks*, 6(1):11.

[413] Zhang, S.; Yang, M.; Cour, T.; Yu, K.; and Metaxas, D. (2015). Query specific rank fusion for image retrieval. *IEEE TPAMI*, 37(4):803–815.

[414] Zhang, T. (2004). Solving large scale linear prediction problems using stochastic gradient descent algorithms. In *Proceedings of the Twenty-First International Conference on Machine Learning*, ICML '04, page 116, New York, NY, USA. Association for Computing Machinery.

[415] Zhang, X.; Ge, Y.; Qiao, Y.; and Li, H. (2021c). Refining pseudo labels with clustering consensus over generations for unsupervised object re-identification. In *2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 3435–3444.

[416] Zhang, Y.; Qian, Q.; Wang, H.; Liu, C.; Chen, W.; and Wang, F. (2024). Graph convolution based efficient re-ranking for visual retrieval. *IEEE Transactions on Multimedia*, 26:1089–1101.

[417] Zhang, Y.; Tian, Y.; Kong, Y.; Zhong, B.; and Fu, Y. (2018b). Residual dense network for image super-resolution. In *CVPR*.

[418] Zhang, Z. and Sabuncu, M. (2018). Generalized cross entropy loss for training deep neural networks with noisy labels. In *Advances in neural information processing systems*, pages 8778–8788.

[419] Zhao, R.; Ouyang, W.; and Wang, X. (2013). Unsupervised salience learning for person re-identification. In *2013 IEEE Conference on Computer Vision and Pattern Recognition*, pages 3586–3593.

[420] Zhao, Y.; Wang, L.; Zhou, L.; Shi, Y.; and Gao, Y. (2018). Modelling diffusion process by deep neural networks for image retrieval. In *British Machine Vision Conference 2018, BMVC 2018, Newcastle, UK, September 3-6, 2018*, page 161. BMVA Press.

[421] Zhao, Z. and Liu, H. (2007). Spectral feature selection for supervised and unsupervised learning. In *Proceedings of the 24th International Conference on Machine Learning*, ICML '07, pages 1151–1157, New York, NY, USA. ACM.

[422] Zheng, L.; Shen, L.; Tian, L.; Wang, S.; Wang, J.; and Tian, Q. (2015a). Scalable person re-identification: A benchmark. In *2015 IEEE International Conference on Computer Vision (ICCV)*, pages 1116–1124.

[423] Zheng, L.; Wang, S.; Liu, Z.; and Tian, Q. (2014a). Packing and padding: Coupled multi-index for accurate image retrieval. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR'2014)*, pages 1947–1954.

[424] Zheng, L.; Wang, S.; Tian, L.; Fei He; Liu, Z.; and Tian, Q. (2015b). Query-adaptive late fusion for image search and person re-identification. In *2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 1741–1750.

[425] Zheng, L.; Wang, S.; and Tian, Q. (2014b). Coupled binary embedding for large-scale image retrieval. *IEEE Transactions on Image Processing (TIP)*, 23(8):3368–3380.

[426] Zheng, L.; Yang, Y.; and Hauptmann, A. G. (2016). Person re-identification: Past, present and future. *CoRR*, abs/1610.02984.

[427] Zheng, L.; Yang, Y.; and Tian, Q. (2018). Sift meets cnn: A decade survey of instance retrieval. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 40(5):1224–1244.

[428] Zheng, Z.; Zheng, L.; and Yang, Y. (2017). Unlabeled samples generated by gan improve the person re-identification baseline in vitro. In *Proceedings of the IEEE International Conference on Computer Vision*, page 3754–3762.

[429] Zhong, Z.; Zheng, L.; Cao, D.; and Li, S. (2017). Re-ranking person re-identification with k-reciprocal encoding. In *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 3652–3661.

[430] Zhong, Z.; Zheng, L.; Li, S.; and Yang, Y. (2018). Generalizing a person retrieval model hetero- and homogeneously. In *The European Conference on Computer Vision (ECCV)*.

[431] Zhong, Z.; Zheng, L.; Luo, Z.; Li, S.; and Yang, Y. (2019). Invariance matters: Exemplar memory for domain adaptive person re-identification. In *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*.

[432] Zhong, Z.; Zheng, L.; Luo, Z.; Li, S.; and Yang, Y. (2020). Learning to adapt invariance in memory for person re-identification. *TPAMI*, pages 1–1.

[433] Zhou, D.; Bousquet, O.; Lal, T. N.; Weston, J.; and Schölkopf, B. (2004). Learning with local and global consistency. In *Advances in Neural Information Processing Systems 16*, pages 321–328. MIT Press.

[434] Zhou, D.; Weston, J.; Gretton, A.; Bousquet, O.; and Schölkopf, B. (2003). Ranking on data manifolds. In *Proceedings of the 16th International Conference on Neural Information Processing Systems*, NIPS'03, page 169–176, Cambridge, MA, USA. MIT Press.

[435] Zhou, K. and Xiang, T. (2019). Torchreid: A library for deep learning person re-identification in pytorch. *arXiv preprint arXiv:1910.10093*.

[436] Zhou, K.; Yang, Y.; Cavallaro, A.; and Xiang, T. (2019a). Learning generalisable omni-scale representations for person re-identification. *arXiv preprint arXiv:1910.06827*.

[437] Zhou, K.; Yang, Y.; Cavallaro, A.; and Xiang, T. (2019b). Omni-scale feature learning for person re-identification. In *The IEEE International Conference on Computer Vision (ICCV)*.

[438] Zhou, W.; Li, H.; and Tian, Q. (2017). Recent advance in content-based image retrieval: A literature survey. *CoRR*, abs/1706.06064.

[439] Zhou, Y. and Croft, W. B. (2006). Ranking robustness: A novel framework to predict query performance. In *ACM Int. Conference on Information and Knowledge Management (CIKM'06)*, pages 567–574.

[440] Zhou, Y. and Croft, W. B. (2007). Query performance prediction in web search environments. In *ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR'07)*, pages 543–550.

[441] Zhou, Y.; Liu, P.; and Qiu, X. (2022). KNN-contrastive learning for out-of-domain intent classification. In Muresan, S.; Nakov, P.; and Villavicencio, A., editors, *Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 5129–5141, Dublin, Ireland. Association for Computational Linguistics.

[442] Zhu, S.; Yang, L.; Chen, C.; Shah, M.; Shen, X.; and Wang, H. (2023). R2former: Unified retrieval and reranking transformer for place recognition. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 19370–19380.

[443] Zhu, X.; Zhu, X.; Li, M.; Morerio, P.; Murino, V.; and Gong, S. (2021). Intra-camera supervised person re-identification. *International Journal of Computer Vision*, 129(5):1580–1595.

[444] Zhu, X.; Zhu, X.; Li, M.; Murino, V.; and Gong, S. (2019). Intra-camera supervised person re-identification: A new benchmark. In *2019 IEEE/CVF International Conference on Computer Vision Workshop (ICCVW)*, pages 1079–1087.

[445] Zoph, B.; Vasudevan, V.; Shlens, J.; and Le, Q. V. (2017). Learning transferable architectures for scalable image recognition. *CoRR*, abs/1707.07012.

[446] Zou, G.; Fu, G.; Peng, X.; Liu, Y.; Gao, M.; and Liu, Z. (2021). Person re-identification based on metric learning: a survey. *Multimedia Tools and Applications*, 80(17):26855–26888.