Writer Identification and Writer Retrieval

Vincent Christlein
Pattern Recognition Lab, Friedrich-Alexander-Universität Erlangen-Nürnberg
SoSe 25

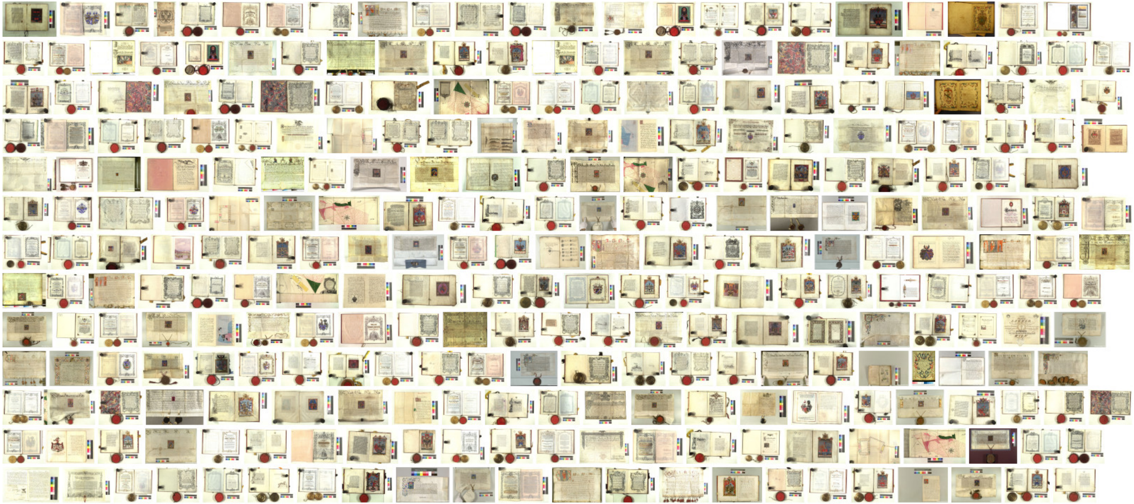
Motivation

Source: Nitramica Arts (CC-BY-SA 2.0), Max Pixel (CC-0)

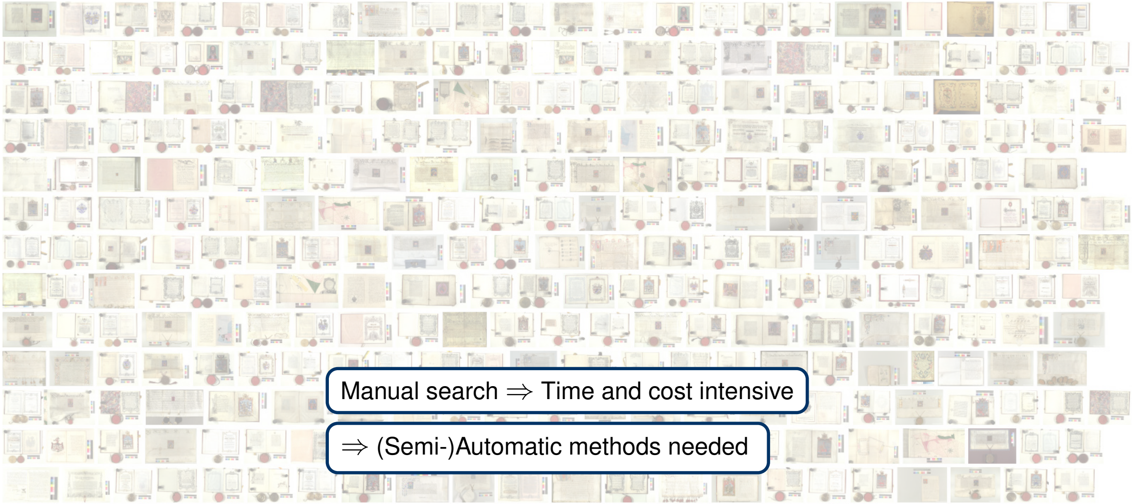
Motivation

Source: Nitramica Arts (CC-BY-SA 2.0), Davide Iliff (CC-BY-SA 3.0)

Handwriting Analysis



Handwriting Analysis



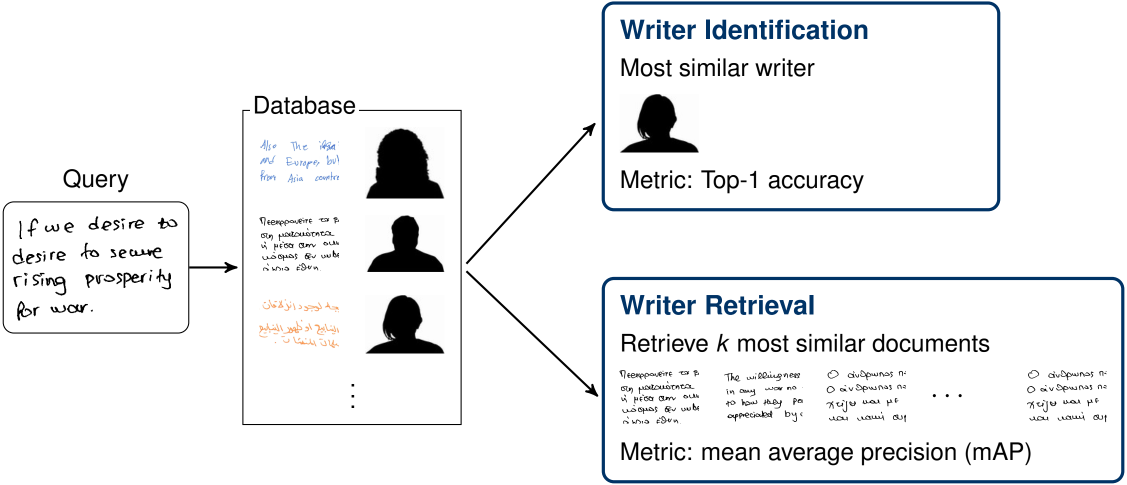
Manual search ⇒ Time and cost intensive

⇒ (Semi-)Automatic methods needed

Outline

Introduction to Writer Identification/Retrieval

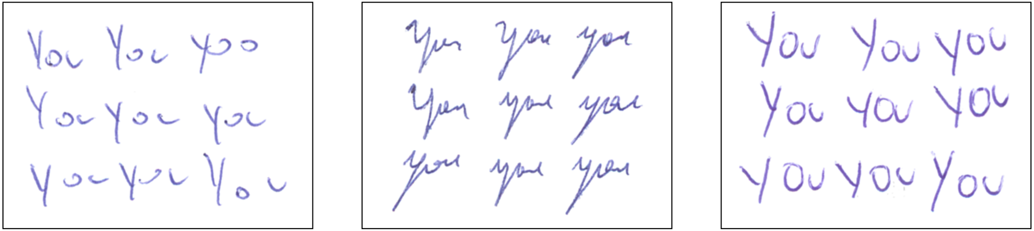
Writer Identification vs. Writer Retrieval



Source: ICDAR13 dataset, QUWI15 dataset, freepik.com

Challenges: Internal Factors

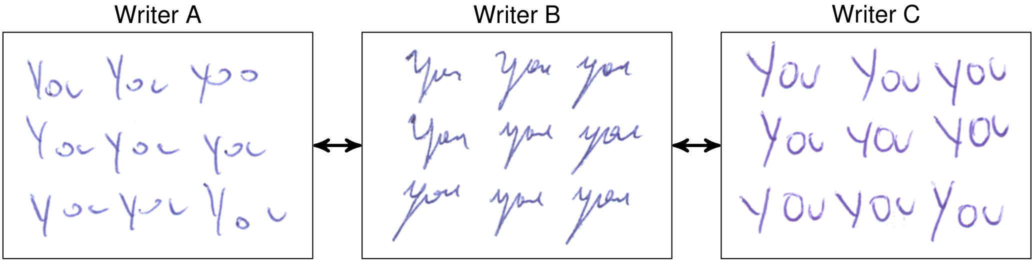
Writer A Writer B Writer C



- Within-writer variability

Source: CVL dataset [1], img-ids (left to right): 0001-7, 0022-7, 0021-7

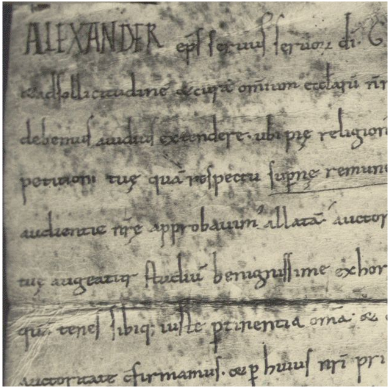
Challenges: Internal Factors

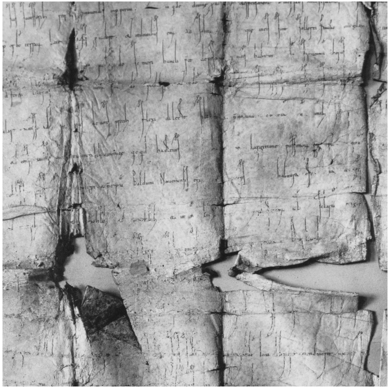


- Within-writer variability
- Between-writer variability

Source: CVL dataset [1], img-ids (left to right): 0001-7, 0022-7, 0021-7

Challenges: External Factors

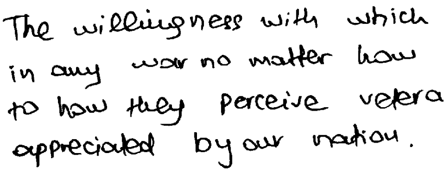


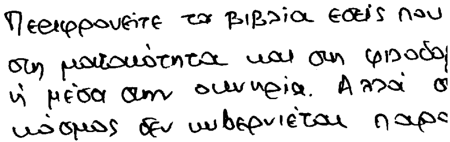


Source:Göttingen Academy of Sciences and Humanities, JL 4490, 4671.

- Pen
- Document Material
- Artifacts

Contemporary Datasets





ICDAR13 benchmark dataset[1]

- 4 documents per writer (2 English, 2 Greek)
- Train: 100 writers → 400 samples
- Test: 250 writers → 1000 samples

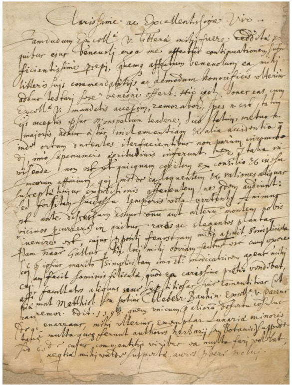
Other datasets: CVL (English, German), KHATT (Arabic), IAM (English)

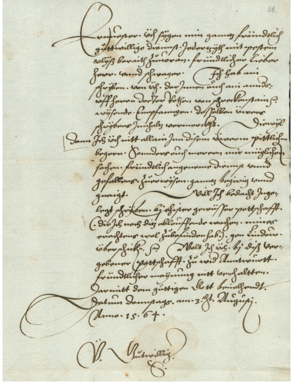
[1] G. Louloudis, B. Gatos, N. Stamatopoulos, and A. Papandreou, "ICDAR 2013 Competition on Writer Identification," in *ICDAR*, Washington DC, NY, Aug. 2013, pp. 1397–1401.

Historical Dataset

ICDAR17 competition dataset

- Letter collection (University library Basel)
- Train: 394 writers x 3 images
  $\rightarrow$ 1182 images
- Test: 720 writers x 5 images
  $\rightarrow$ 3600 images

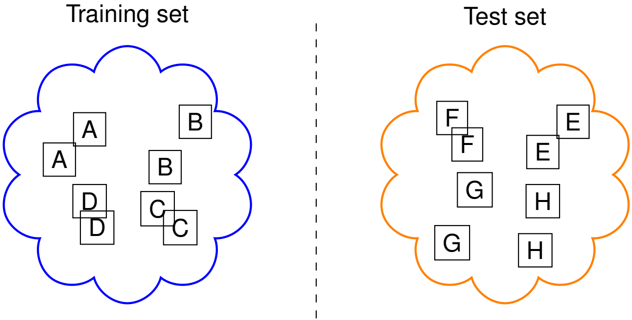




Source: ICDAR17 Historical-WI, ID: 2056-IMG_MAX_320331, 1146-3-IMG_MAX_1207684

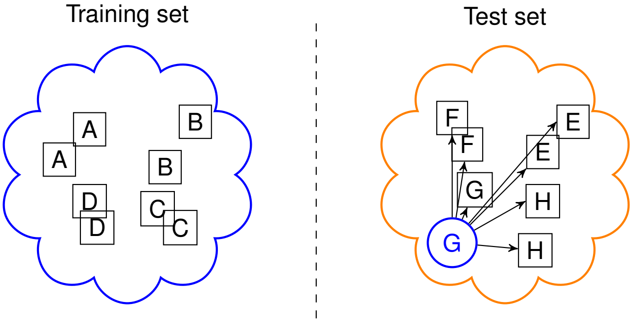
S. Fiel, F. Kleber, M. Diem, V. Christlein, G. Louloudis, N. Stamatopoulos, and B. Gatos, "ICDAR2017 Competition on Historical Document Writer Identification," in *ICDAR*, 2013

Writer-Independent Datasets



Training and test sets are independent
$\Rightarrow$ No training for a specific writer possible!

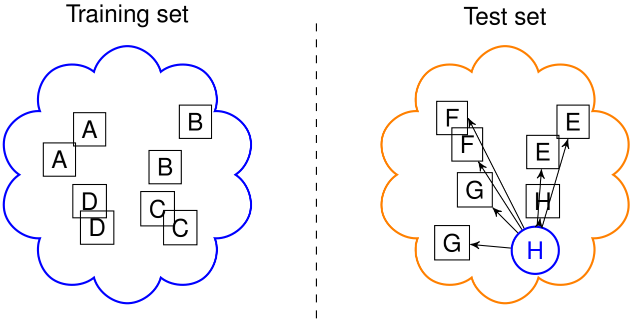
Writer-Independent Datasets



Training and test sets are independent

$\Rightarrow$ No training for a specific writer possible!

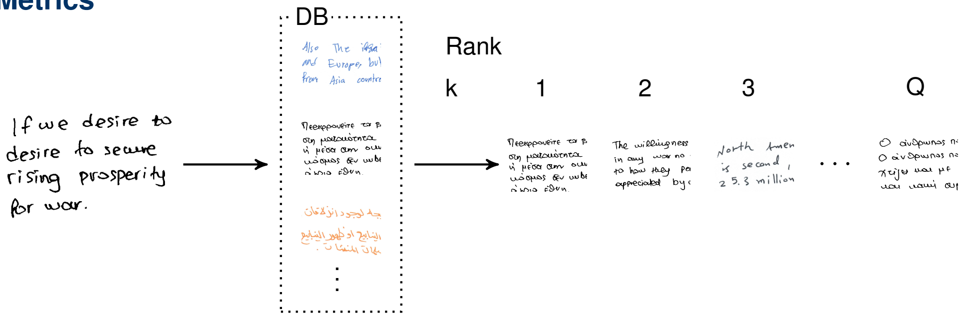
Writer-Independent Datasets



Training and test sets are independent

$\Rightarrow$ No training for a specific writer possible!

Error Metrics



Identification rate

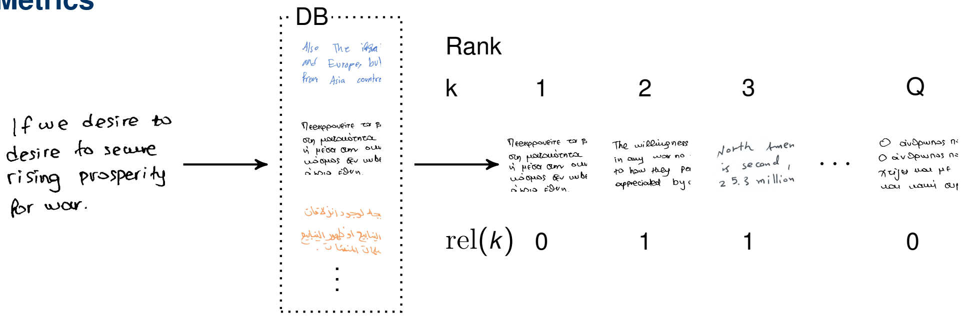
Mean precision at rank 1

Mean average precision

$$\text{mAP} = \frac{1}{N}\sum_{i=1}^{N} \text{AP}_i \quad N\text{: \#queries}$$

Source: ICDAR13 dataset, QUWI15 dataset

Error Metrics



Identification rate

Mean precision at rank 1

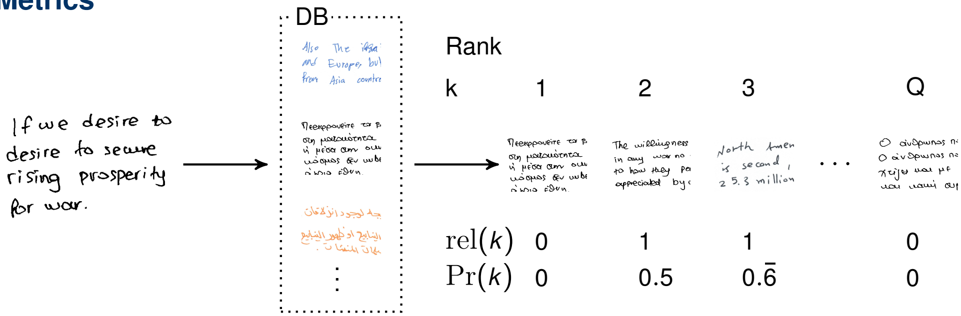
Mean average precision

$$\mathrm{mAP} = \frac{1}{N}\sum_{i=1}^{N} \mathrm{AP}_i \quad N\text{: \#queries}$$

$$\mathrm{AP}_i = \frac{\sum_{k=1}^{Q} \Pr(k)\cdot \mathrm{rel}(k)}{\text{number of relevant documents}}$$

Source: ICDAR13 dataset, QUWI15 dataset

Error Metrics



k	1	2	3	…	Q
$\mathrm{rel}(k)$	0	1	1		0
$\Pr(k)$	0	0.5	$0.\overline{6}$		0

Identification rate

Mean precision at rank 1

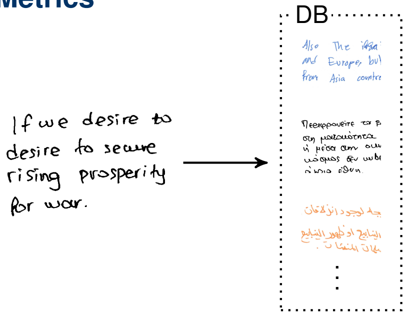
Mean average precision

$$\mathrm{mAP} = \frac{1}{N}\sum_{i=1}^{N} \mathrm{AP}_i$$ N: #queries

$$\mathrm{AP}_i = \frac{\sum_{k=1}^{Q} \Pr(k)\cdot \mathrm{rel}(k)}{\text{number of relevant documents}}$$

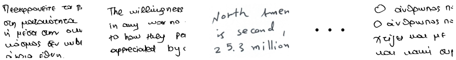
Source: ICDAR13 dataset, QUWI15 dataset

Error Metrics



Rank

k	1	2	3	…	Q



$\mathrm{rel}(k)$	0	1	1		0
$\mathrm{Pr}(k)$	0	0.5	$0.\bar{6}$		0

$$\mathrm{AP} = (0.5 + 0.\bar{6})/2 \approx 0.58$$

Identification rate

Mean precision at rank 1

Mean average precision

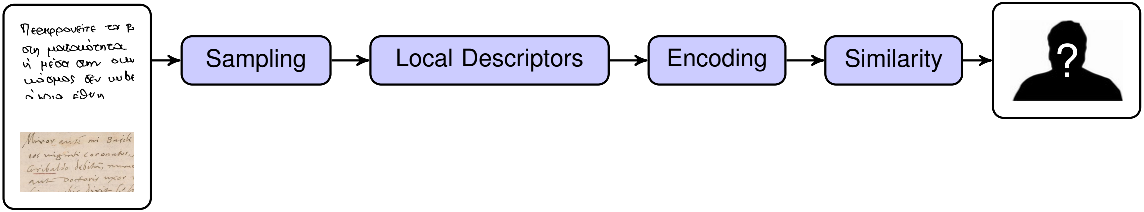
$$\mathrm{mAP} = \frac{1}{N}\sum_{i=1}^{N} \mathrm{AP}_i \quad N\text{: \#queries}$$

$$\mathrm{AP}_i = \frac{\sum_{k=1}^{Q} \mathrm{Pr}(k)\cdot\mathrm{rel}(k)}{\text{number of relevant documents}}$$

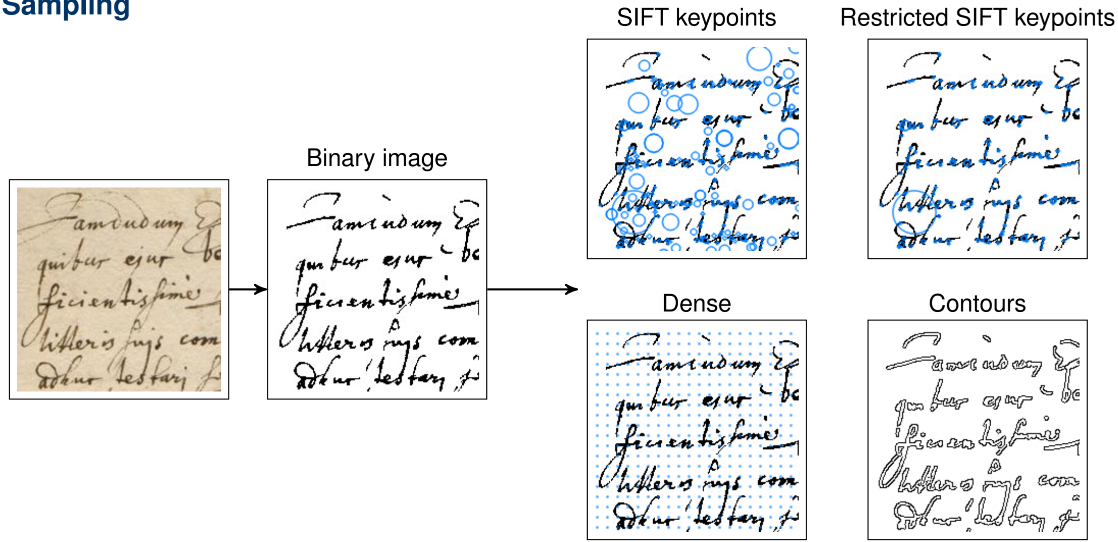
Source: ICDAR13 dataset, QUWI15 dataset

General Approach

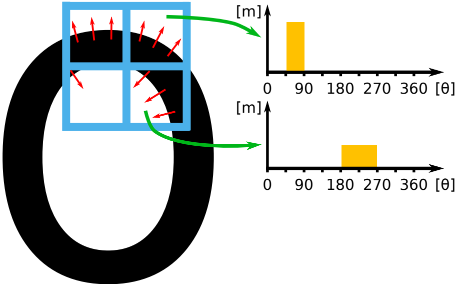
Methodology



Sampling


Binary image
SIFT keypoints
Restricted SIFT keypoints
Dense
Contours

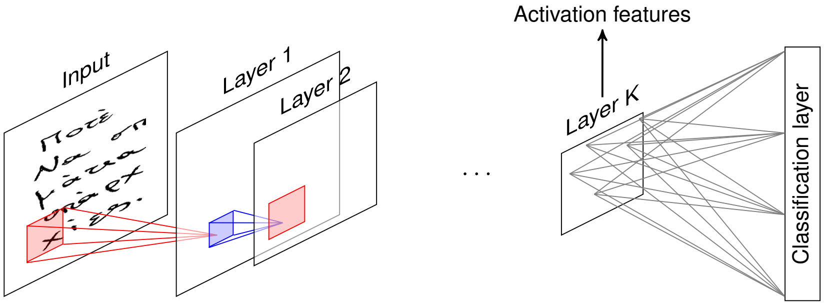
Handcrafted Features



SIFT: Scale-Invariant Feature Transform[2]

[2] D. G. Lowe, “Distinctive Image Features from Scale-Invariant Keypoints,” *International Journal of Computer Vision*, vol. 60, no. 2, pp. 91–110, Nov. 2004

Convolutional Neural Network Activation Features (CNN AF)



- Recall: no end-to-end training possible ⇒ one-shot learning

⇒ Surrogate task: classify writers of the training set using cross-entropy ("soft-max loss")

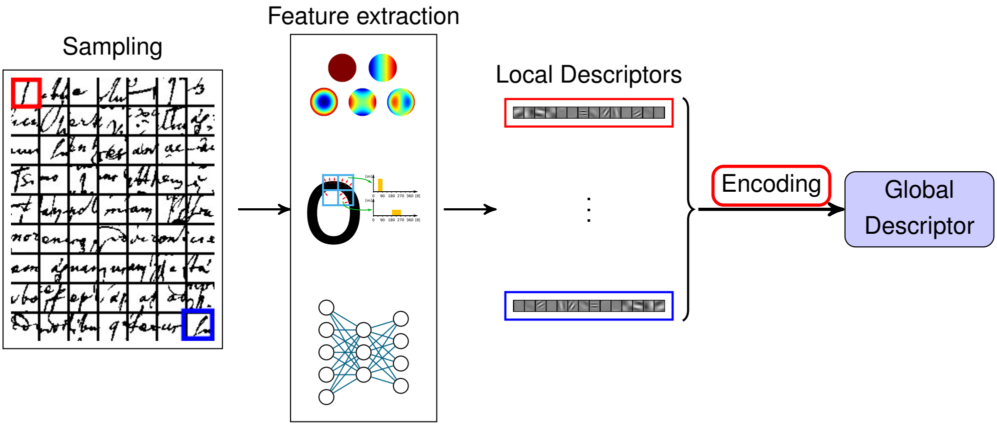
- Use CNN as feature extractor

Metric Learning-based Features

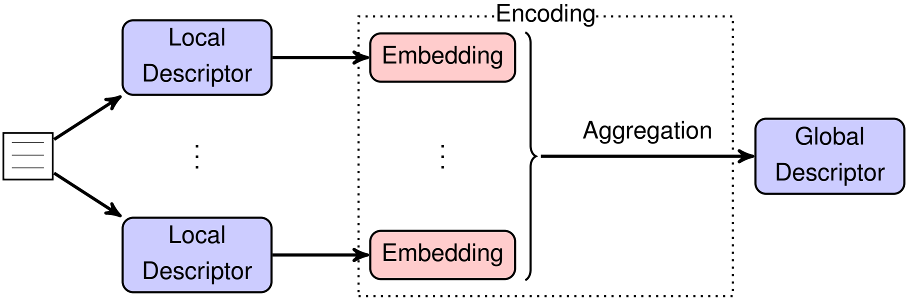
- Contrastive Loss
- Triplet Loss
- Magnet Loss
- Histogram Loss
- ...

Unsupervised: e.g. AutoEncoders

Global Representation



Encoding

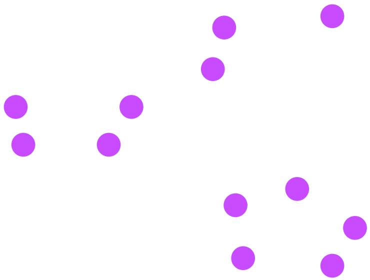


- Embedding: Map into high dimensional representation
- Aggregation: Sum pooling, generalized max-pooling[3]
- Normalization + Decorrelation

[3] N. Murray, H. Jegou, F. Perronnin, and A. Zisserman, "Interferences in Match Kernels," *TPAMI*, vol. 39, no. 9, 2016.

VLAD Embedding

VLAD: Vectors of Locally Aggregated Descriptors[4]

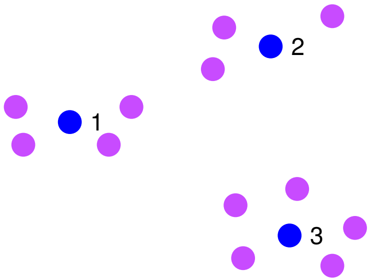


Local Descriptors: $\mathcal{X} = \{\boldsymbol{x}_i \in \mathbb{R}^D, i = 1, \ldots, T\}$

[4] H. Jégou, F. Perronnin, M. Douze, J. Sánchez, P. Pérez, and C. Schmid, "Aggregating Local Image Descriptors into Compact Codes.," *PAMI*, vol. 34, no. 9, 2012.

VLAD Embedding

VLAD: Vectors of Locally Aggregated Descriptors[4]



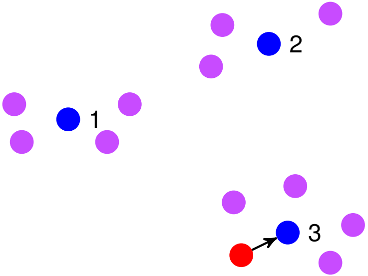
Local Descriptors: $\mathcal{X} = \{\boldsymbol{x}_i \in \mathbb{R}^D, i = 1, \dots, T\}$

Clusters: $\mathcal{D} = \{\boldsymbol{\mu}_k \in \mathbb{R}^D, k = 1, \dots, K\}$

[4] H. Jégou, F. Perronnin, M. Douze, J. Sánchez, P. Pérez, and C. Schmid, "Aggregating Local Image Descriptors into Compact Codes.," *PAMI*, vol. 34, no. 9, 2012.

VLAD Embedding

VLAD: Vectors of Locally Aggregated Descriptors[4]



Local Descriptors: $\mathcal{X} = \{\boldsymbol{x}_i \in \mathbb{R}^D, i = 1, \dots, T\}$

Clusters: $\mathcal{D} = \{\boldsymbol{\mu}_k \in \mathbb{R}^D, k = 1, \dots, K\}$

$$\phi_k(\boldsymbol{x}) = \alpha_k(\boldsymbol{x})(\boldsymbol{x} - \boldsymbol{\mu}_k)$$

$$\alpha_k(\boldsymbol{x}) = \begin{cases} 1 & \text{if } k = \underset{j=1,\dots,K}{\operatorname{argmin}} \|\boldsymbol{x} - \boldsymbol{\mu}_j\|_2 \\ 0 & \text{else} \end{cases}$$

$$\phi_{\text{VLAD}}(\boldsymbol{x}) = \left(\phi_1^\top, \dots, \phi_K^\top\right)^\top \in \mathbb{R}^{D \cdot K}$$

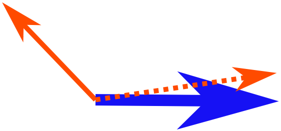
$$\phi_{\text{VLAD}}(\boldsymbol{x}) = (\overbrace{\mathbf{0}}^{\in \mathbb{R}^D}, \overbrace{\mathbf{0}}^{\in \mathbb{R}^D}, \overbrace{(\boldsymbol{x} - \boldsymbol{\mu}_3)^\top}^{\in \mathbb{R}^D})^\top$$

[4] H. Jégou, F. Perronnin, M. Douze, J. Sánchez, P. Pérez, and C. Schmid, "Aggregating Local Image Descriptors into Compact Codes.," *PAMI*, vol. 34, no. 9, 2012.

Sum Pooling vs. Generalized Max Pooling

Visual Burstiness

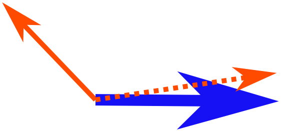
Sum pooling



- Unrelated descriptors produce interference
- Frequent descriptors dominate similarity

Visual Burstiness

Sum pooling



- Unrelated descriptors produce interference
- Frequent descriptors dominate similarity

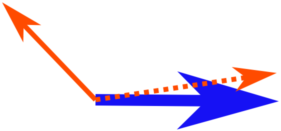
→ Choose better embedding

→ Normalize encoding

  - Power normalization
  - Intra normalization
  - ...

Visual Burstiness

Sum pooling

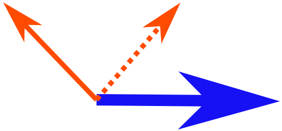


- Unrelated descriptors produce interference
- Frequent descriptors dominate similarity

➔ Choose better embedding

➔ Normalize encoding
  - Power normalization
  - Intra normalization
  - ...

Generalized max pooling [5]



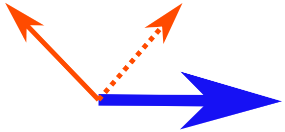
➔ Balance pooling

Generalized Max Pooling

- Seek encoding ξ which weights each embedding ϕ

$$\boldsymbol{\xi} = \sum_{\boldsymbol{x} \in \mathcal{X}} \alpha(\boldsymbol{x}) \phi(\boldsymbol{x}) = \boldsymbol{\alpha}^T \boldsymbol{\Phi}$$

Generalized max pooling [5]



Generalized Max Pooling

- Seek encoding ξ which weights each embedding ϕ

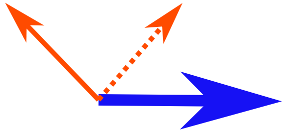
$$\xi = \sum_{\boldsymbol{x} \in \mathcal{X}} \alpha(\boldsymbol{x})\phi(\boldsymbol{x}) = \boldsymbol{\alpha}^T \boldsymbol{\Phi}$$

- Max pooling: equally similar to frequent and rare patches

→ Enforce similarity between any patch encoding and aggregated representation to be constant

$$\boldsymbol{\Phi}^\top \xi_{\text{gmp}} = \mathbf{1}_n,$$

Generalized max pooling [5]



Generalized Max Pooling

- Seek encoding ξ which weights each embedding ϕ

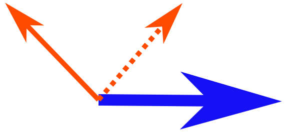
$$\xi = \sum_{\boldsymbol{x} \in \mathcal{X}} \alpha(\boldsymbol{x})\phi(\boldsymbol{x}) = \boldsymbol{\alpha}^T \boldsymbol{\Phi}$$

- Max pooling: equally similar to frequent and rare patches

→ Enforce similarity between any patch encoding and aggregated representation to be constant

$$\boldsymbol{\Phi}^\top \xi_{\text{gmp}} = \mathbf{1}_n,$$

Generalized max pooling [5]



→ Optimization problem can be cast as a ridge regression problem

$$\xi_{\text{gmp}} = \underset{\xi}{\operatorname{argmin}} \|\boldsymbol{\Phi}^\top \xi - \mathbf{1}_n\|^2 + \lambda \|\xi\|^2 \,,$$

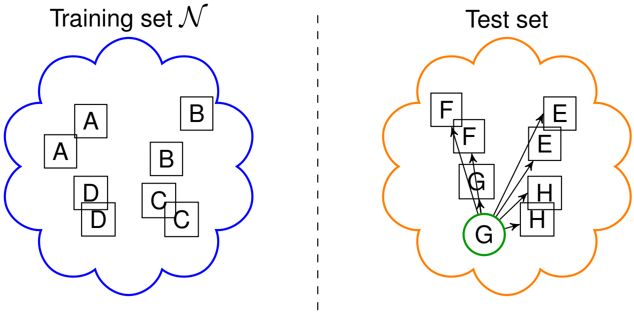
$\lambda \to 0$: max pooling

$\lambda \to \infty$: sum pooling

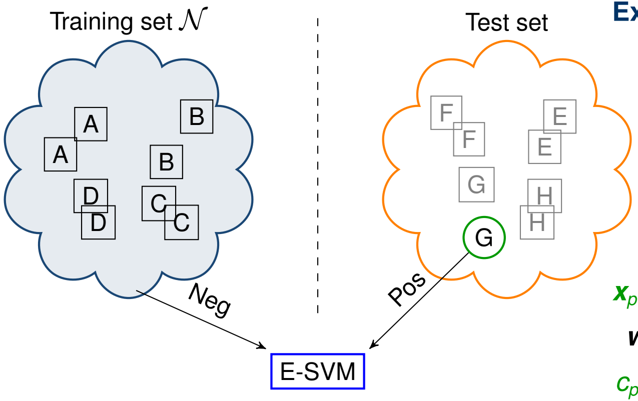
FRIEDRICH-ALEXANDER-UNIVERSITÄT ERLANGEN-NÜRNBERG
SCHOOL OF ENGINEERING

Exemplar Classification

Similarity



Similarity



$\Rightarrow$ Subject-specific similarity!

Exemplar SVMs [Christlein17a]

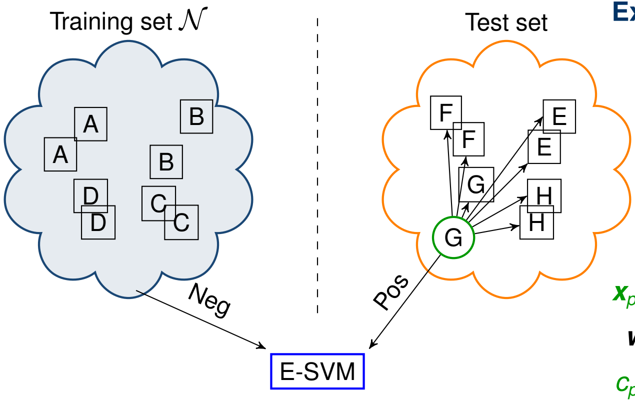
$$\min_{\boldsymbol{w},b} \frac{1}{2}\|\boldsymbol{w}\|^2 + c_p \max(0, 1 - \boldsymbol{w}^\top \boldsymbol{x}_p - b)^2 + c_n \sum_{\boldsymbol{x}_n \in \mathcal{N}} \max(0, 1 + \boldsymbol{w}^\top \boldsymbol{x}_n + b)^2$$

$\boldsymbol{x}_p, \boldsymbol{x}_n$: query sample, background sample

$\boldsymbol{w}, b$: model parameters

c_p, c_n : margin parameters (e.g. indirect proportional to #samples)

Similarity



$\Rightarrow$ Subject-specific similarity!

Exemplar SVMs [Christlein17a]

$$\min_{\boldsymbol{w},b} \frac{1}{2}\|\boldsymbol{w}\|^2 + c_p \max(0, 1 - \boldsymbol{w}^\top \boldsymbol{x}_p - b)^2 + c_n \sum_{\boldsymbol{x}_n \in \mathcal{N}} \max(0, 1 + \boldsymbol{w}^\top \boldsymbol{x}_n + b)^2$$

$\boldsymbol{x}_p, \boldsymbol{x}_n$: query sample, background sample

$\boldsymbol{w}, b$: model parameters

c_p, c_n : margin parameters (e.g. indirect proportional to #samples)

Exemplar SVMs as Feature Encoder

- Trained SVM models: $\boldsymbol{w}$, b
- Similarity independent of b when using cosine distance

$\Rightarrow$ New feature: $\boldsymbol{x} \rightarrow \frac{\boldsymbol{w}}{\|\boldsymbol{w}\|_2}$

- Note: iterative application of E-SVM possible but benefit vanishes quickly

Thank you for your attention

? ?

Questions? Questions? Questions? Questions? Questions? Questions?

Questions? Questions? Questions? Questions? Questions? Questions?

Questions? Questions? Questions? Questions? Questions? Questions?

References

References I

[1] F. Kleber, S. Fiel, M. Diem, and R. Sablatnig, “CVL-DataBase: An Off-Line Database for Writer Retrieval, Writer Identification and Word Spotting,” in *Document Analysis and Recognition (ICDAR), 2013 12th International Conference on*, Washington DC, NY, Aug. 2013, pp. 560–564.

[2] G. Louloudis, B. Gatos, N. Stamatopoulos, and A. Papandreou, “ICDAR 2013 Competition on Writer Identification,” in *ICDAR*, Washington DC, NY, Aug. 2013, pp. 1397–1401.

[3] S. Fiel, F. Kleber, M. Diem, V. Christlein, G. Louloudis, N. Stamatopoulos, and B. Gatos, “ICDAR2017 Competition on Historical Document Writer Identification,” in *ICDAR*, 2013.

[4] D. G. Lowe, “Distinctive Image Features from Scale-Invariant Keypoints,” *International Journal of Computer Vision*, vol. 60, no. 2, pp. 91–110, Nov. 2004.

[5] N. Murray, H. Jegou, F. Perronnin, and A. Zisserman, “Interferences in Match Kernels,” *TPAMI*, vol. 39, no. 9, 2016.

[6] H. Jégou, F. Perronnin, M. Douze, J. Sánchez, P. Pérez, and C. Schmid, “Aggregating Local Image Descriptors into Compact Codes.,” *PAMI*, vol. 34, no. 9, 2012.