



**FAU**

FRIEDRICH-ALEXANDER-  
UNIVERSITÄT  
ERLANGEN-NÜRNBERG  
SCHOOL OF ENGINEERING

# 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 Iliiff (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**

**General Approach**

**Sum Pooling vs. Generalized Max Pooling**

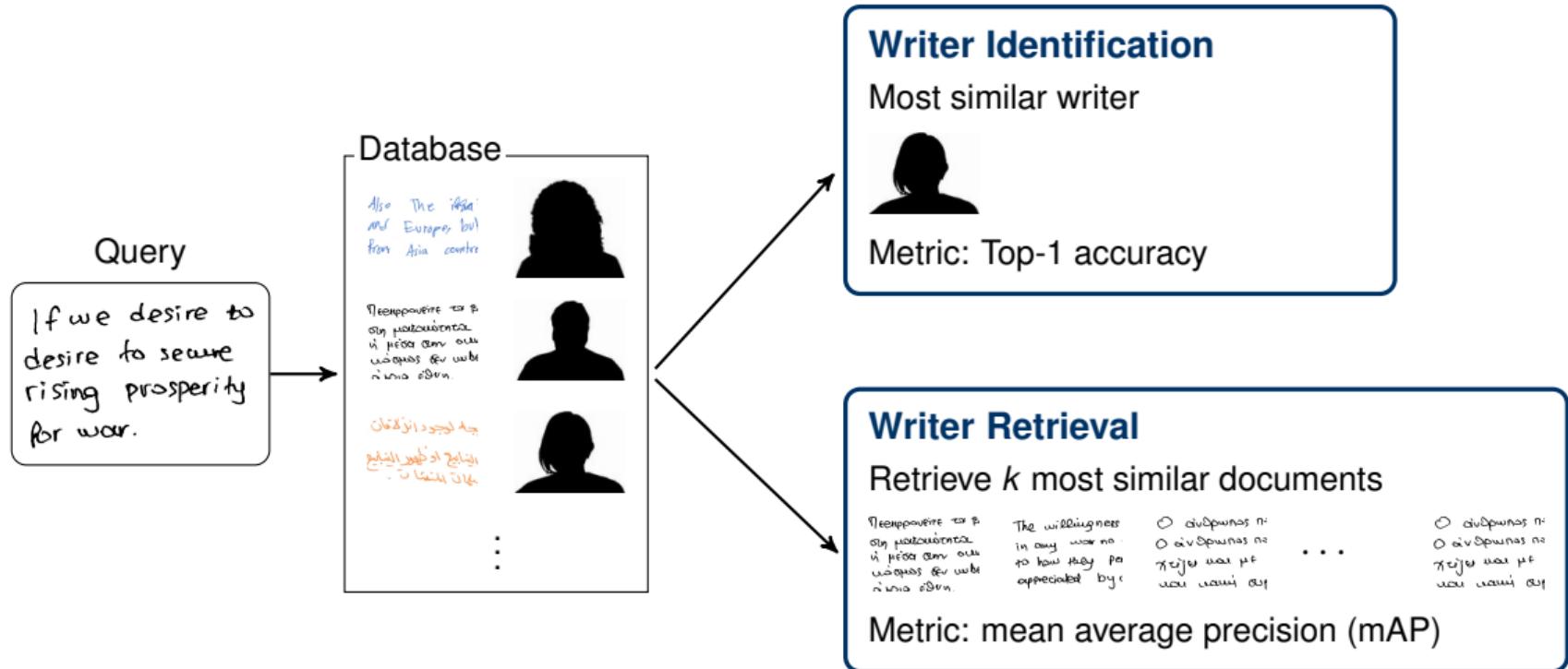
**Exemplar Classification**



# Introduction to Writer Identification/Retrieval



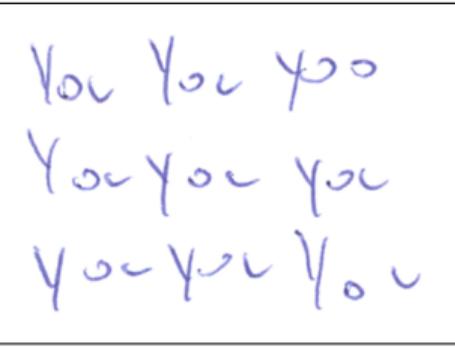
## Writer Identification vs. Writer Retrieval



Source: ICDAR13 dataset, QUWI15 dataset, freepik.com

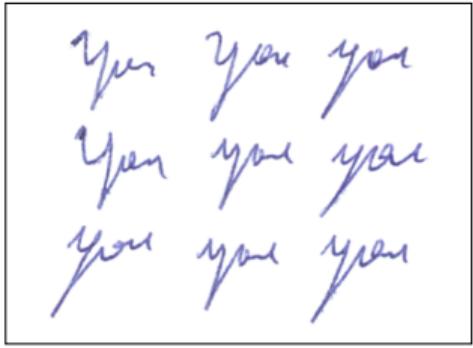
## Challenges: Internal Factors

Writer A



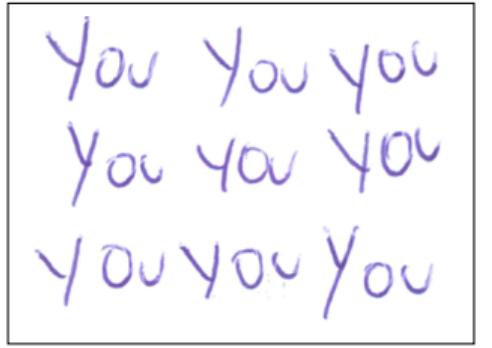
You You You  
You You You  
You You You

Writer B



Yer Yer Yer  
Yer Yer Yer  
Yer Yer Yer

Writer C

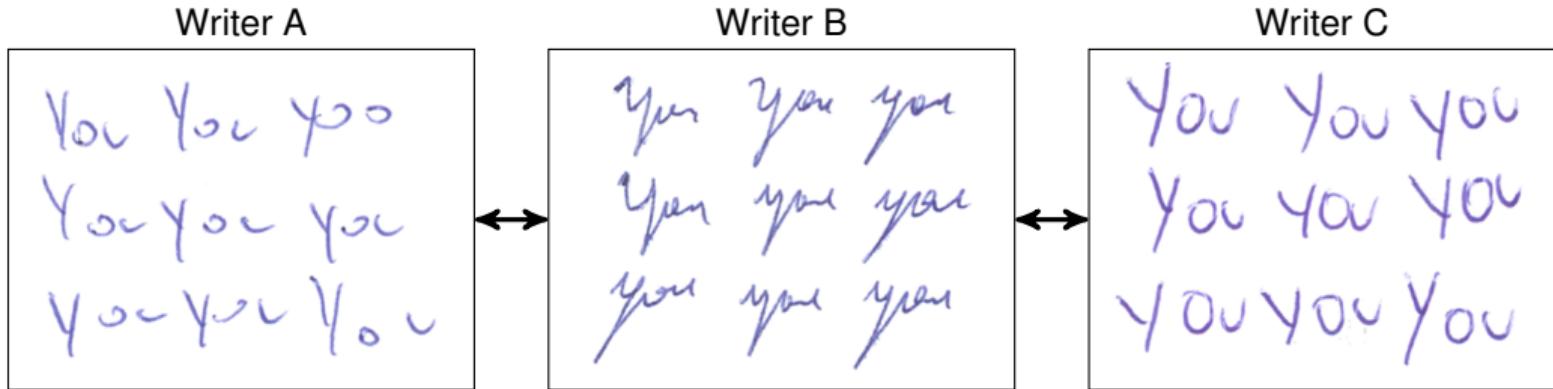


You You You  
You You You  
You You You

- 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

ALEXANDER ep[iscopus] forus seruor[um] d[omi]ni G[loria] C[oncordia]  
ad collaudatione decimae omium eccl[esi]aq[ue] nr[um]  
debetus audire excedere ubi p[ro]p[ter] religionis  
petitione tue qua respectu supne remun  
audiencie tue approbaum illata auctor  
tue augearit flaciu[m] benignissime exhor  
qui tenet sibiq[ue] iuste pertinencia omnia se  
succurrat firmamus. et p[ro]p[ter] huius nr[um] pri



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

- Pen
- Document Material
- Artifacts

## Contemporary Datasets

The willingness with which  
in any war no matter how  
to how they perceive veterans  
appreciated by our nation.

Πεπονωτε τα βίβλια τούς να  
συμμαχήσει με την ψηφοδέλτιο  
η μέσα στην συνείδηση. Απότι ο  
νικητής δεν ωθεψείται πάρα

### ICDAR13 benchmark dataset<sup>1</sup>

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

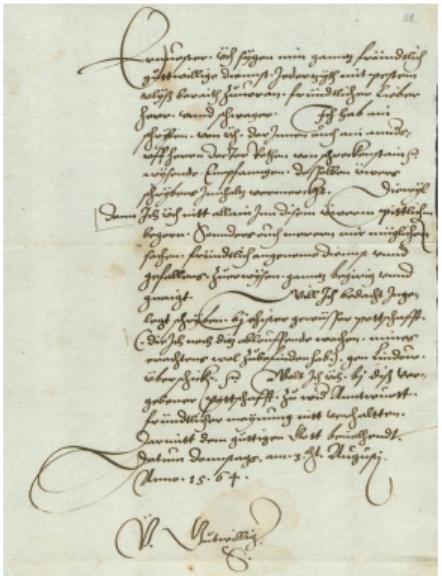
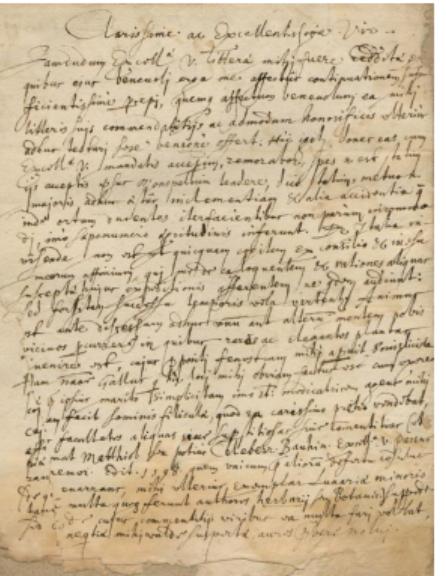
Other datasets: CVL (English, German), KHATT (Arabic), IAM (English)

<sup>1</sup>G. Louloudis, B. Galos, 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

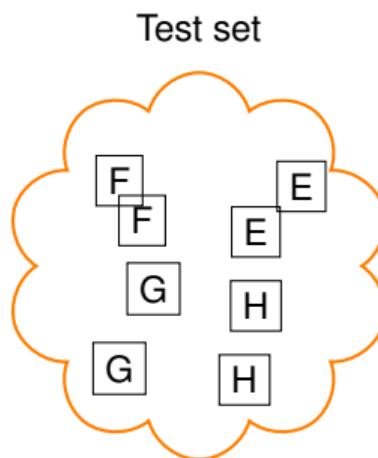
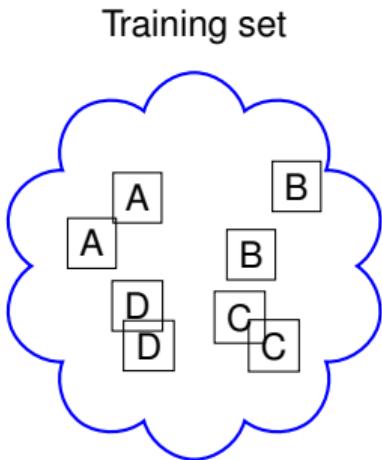
- Letter collection (University library Basel)
- Train: 394 writers x 3 images  
→ 1182 images
- Test: 720 writers x 5 images  
→ 3600 images



Source: ICDAR17 Historical-WI, ID: 2056-IMG\_MAX\_320331, 1146-3-IMG\_MAX\_1207684

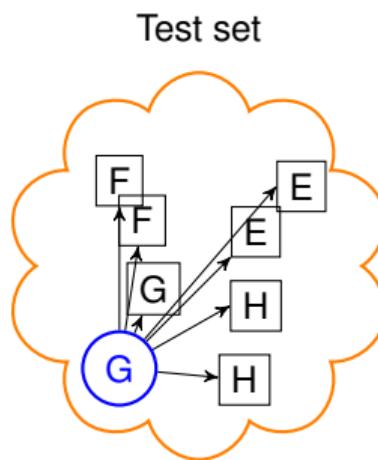
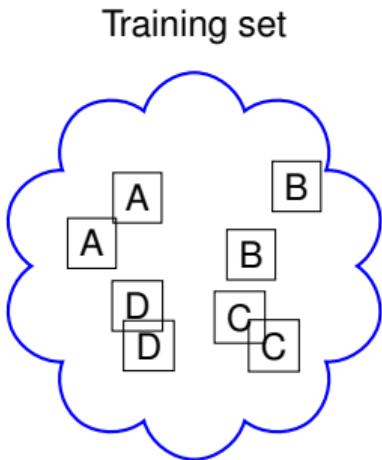
S. Fiel, F. Kleber, M. Diem, V. Christlein, G. Louloudis, N. Stamatopoulos, and B. Galos, "ICDAR2017 Competition on Historical Document Writer Identification," in *ICDAR*, 2013

## Writer-Independent Datasets



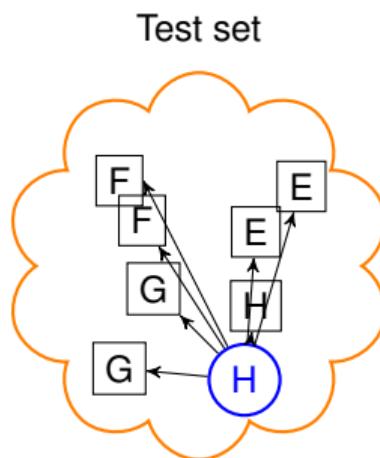
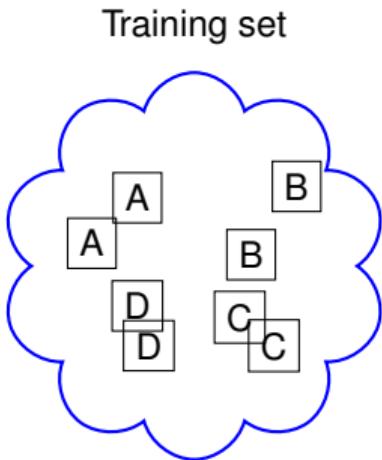
Training and test sets are independent  
⇒ No training for a specific writer possible!

## Writer-Independent Datasets



Training and test sets are independent  
⇒ No training for a specific writer possible!

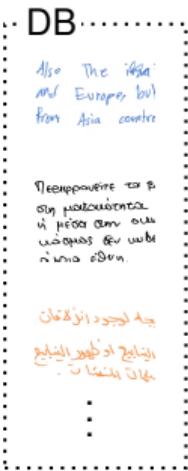
## Writer-Independent Datasets



Training and test sets are independent  
⇒ No training for a specific writer possible!

## Error Metrics

If we desire to  
desire to secure  
rising prosperity  
for war.



### Rank

k 1

2

3

Q

The willingness  
to be  
influence  
is based on our  
wishes for who  
is more often

The willingness  
in any war no  
to how they are  
appreciated by

North Amer  
is second,  
2.5 million

...

O disappears in  
O disappears no  
they are not  
seen want day

## Identification rate

Mean precision at rank 1

## Mean average precision

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

## Error Metrics

If we desire to  
desire to secure  
rising prosperity  
for war.



### Rank

k 1

2

3

Q

rel( $k$ ) 0

1

1

0

The willingness  
in any war no.  
to how they be  
appreciated by

North Amer  
is second,  
2.5.3 million

⋮  
○ disappears in  
○ disappears in  
they war pt  
you want day

## Identification rate

Mean precision at rank 1

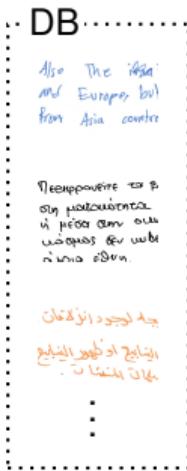
## Mean average precision

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

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

## Error Metrics

If we desire to  
desire to secure  
rising prosperity  
for war.



### Rank

k 1

2

3

Q

Neckermann to be our most important in press over our willingness for war in Asia often

The willingness in any war no to how they for appreciated by

North Amer is second, 2.5.3 million

...  
O oispoonas n  
O oispoonas n  
Tizje war pf  
you want day

rel( $k$ ) 0

1 1

0

Pr( $k$ ) 0

0.5 0.6

0

## Identification rate

Mean precision at rank 1

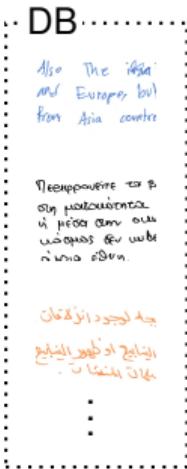
## Mean average precision

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

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

## Error Metrics

If we desire to  
desire to secure  
rising prosperity  
for war.



### Rank

k	1	2	3	Q
---	---	---	---	---

→

Relevant to 3 of 10 documents in press can our answers for war in Asia center	The willingness in any war no. to how they are appreciated by	North Amer is second, 2.5.3 million	...	○ documents in ○ documents no true war pt can want day
---	---	-------------------------------------	-----	---

rel( $k$ )	0	1	1	0
Pr( $k$ )	0	0.5	0.6	0

$$AP = (0.5 + 0.6)/2 \approx 0.58$$

## Identification rate

Mean precision at rank 1

## Mean average precision

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

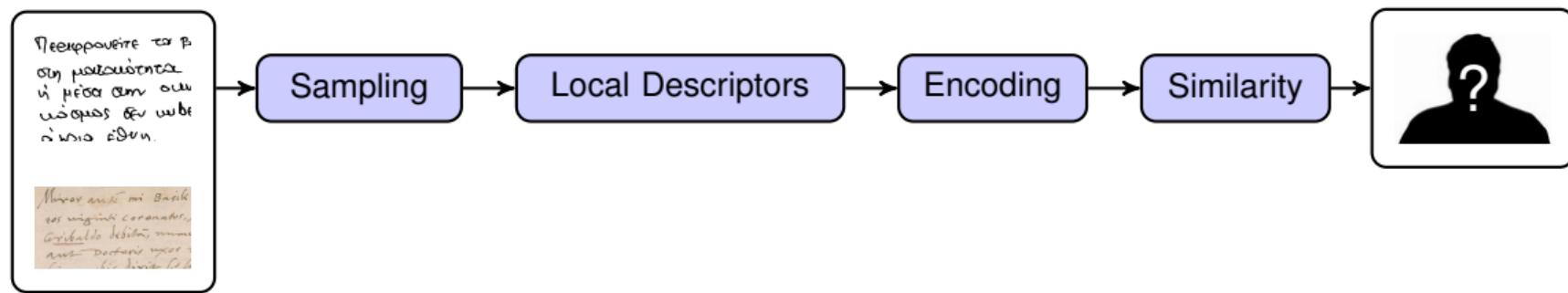
$$AP_i = \frac{\sum_{k=1}^Q Pr(k) \cdot rel(k)}{\text{number of relevant documents}}$$



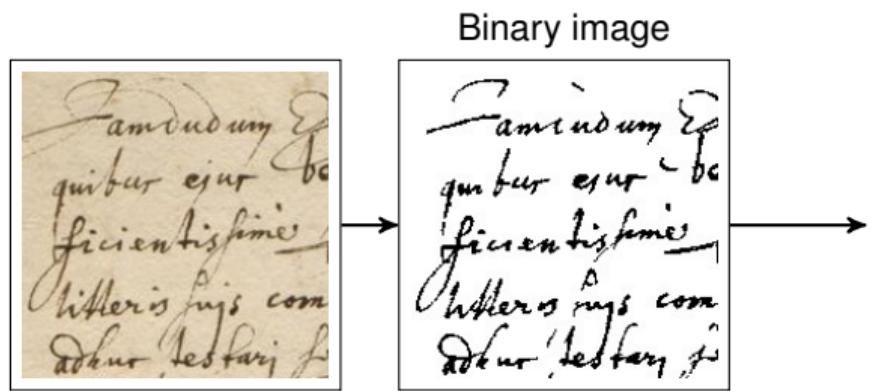
# General Approach



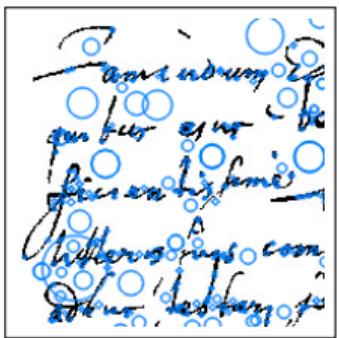
# Methodology



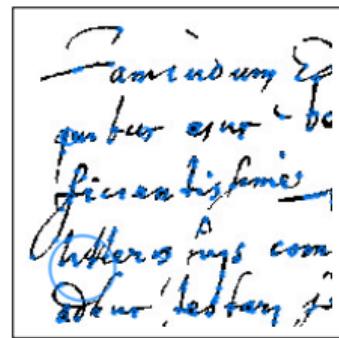
# Sampling



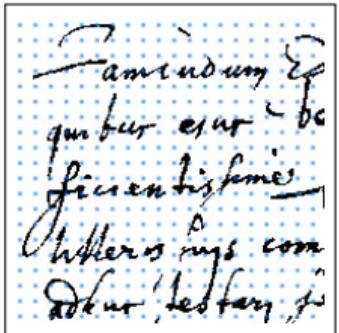
SIFT keypoints



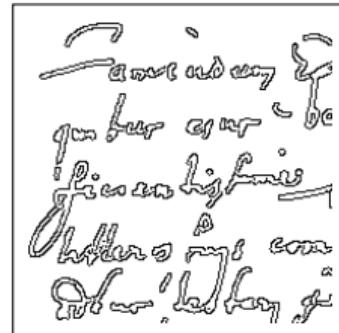
Restricted SIFT keypoints



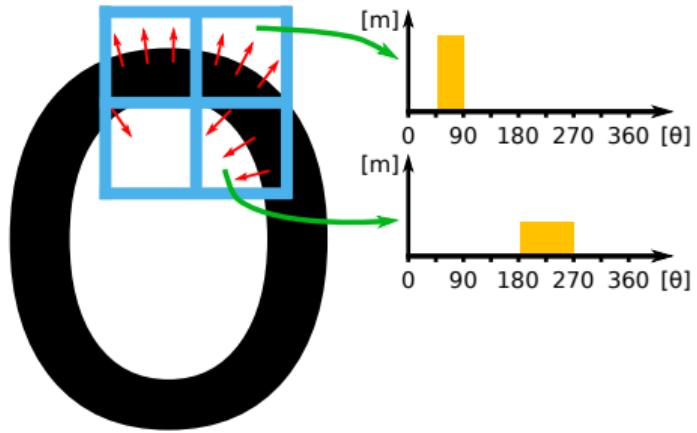
Dense



Contours



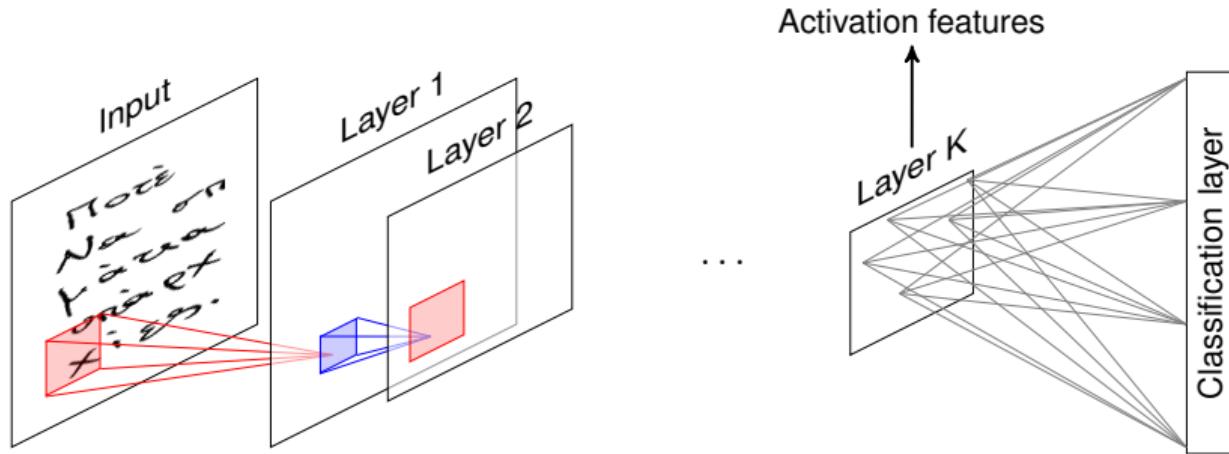
## Handcrafted Features



SIFT: Scale-Invariant Feature Transform<sup>2</sup>

<sup>2</sup>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)



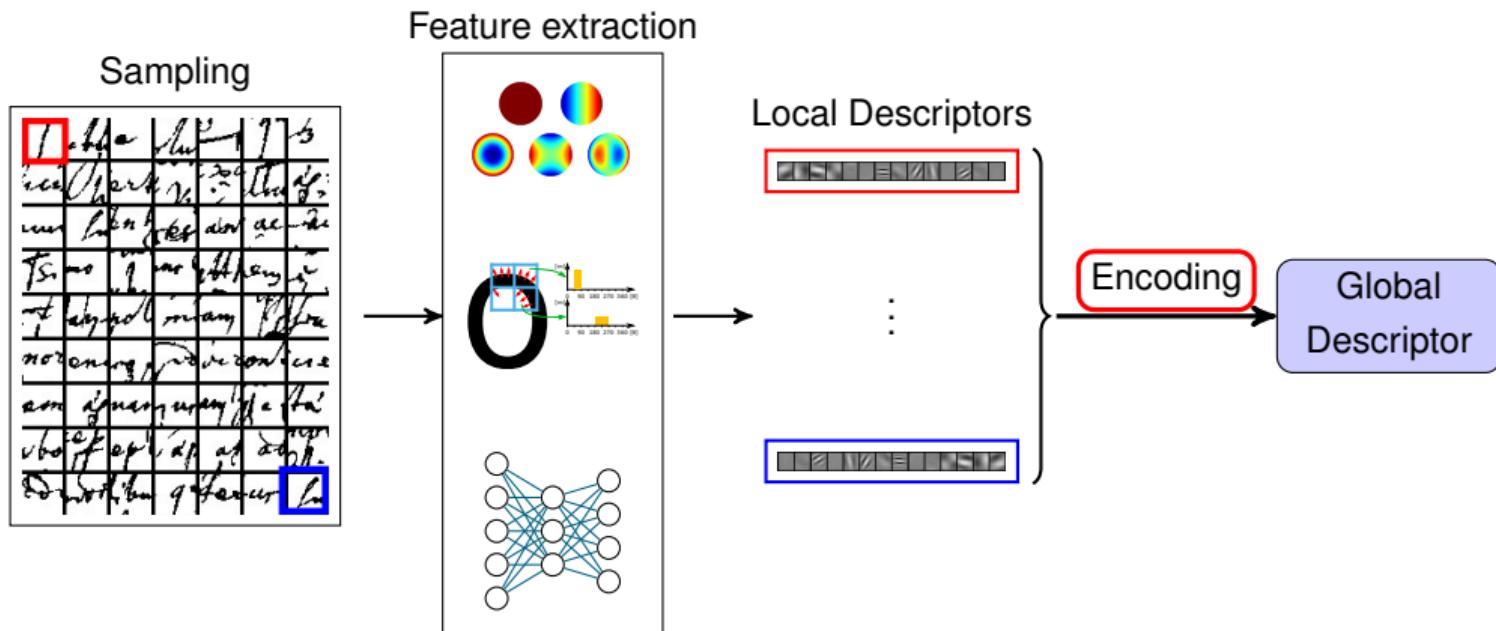
- Recall: no end-to-end training possible  $\Rightarrow$  one-shot learning
- $\Rightarrow$  Surrogate task: classify writers of the training set using cross-entropy ("soft-max loss")
- Use CNN as feature extractor

## Metric Learning-based Features

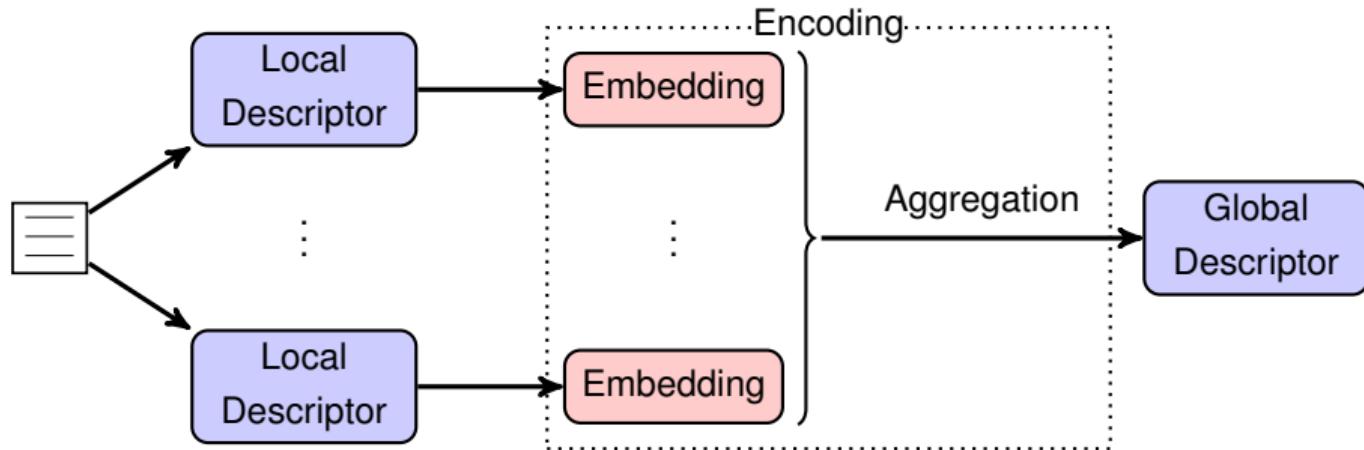
- Contrastive Loss
- Triplet Loss
- Magnet Loss
- Histogram Loss
- ...

**Unsupervised: e.g. AutoEncoders**

# Global Representation



## Encoding

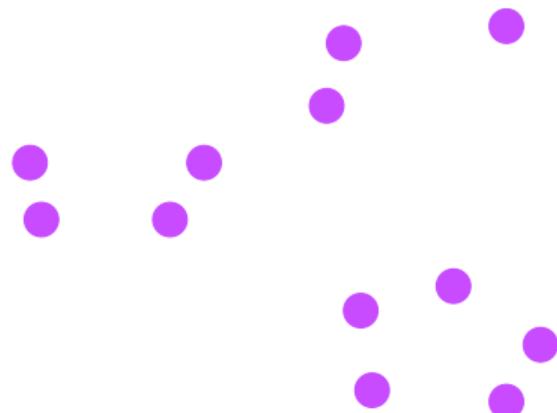


- Embedding: Map into high dimensional representation
- Aggregation: Sum pooling, generalized max-pooling<sup>3</sup>
- Normalization + Decorrelation

<sup>3</sup>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<sup>4</sup>



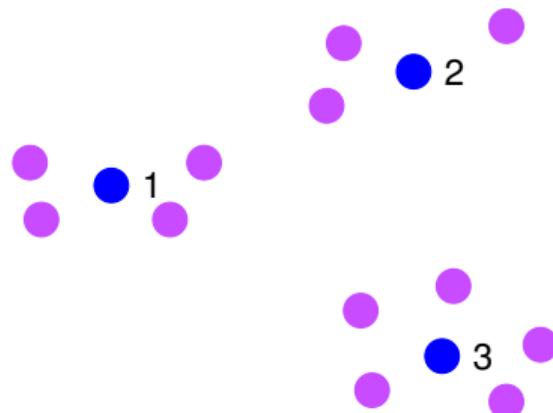
Local Descriptors:  $\mathcal{X} = \{\mathbf{x}_i \in \mathbb{R}^D, i = 1, \dots, T\}$

---

<sup>4</sup>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<sup>4</sup>

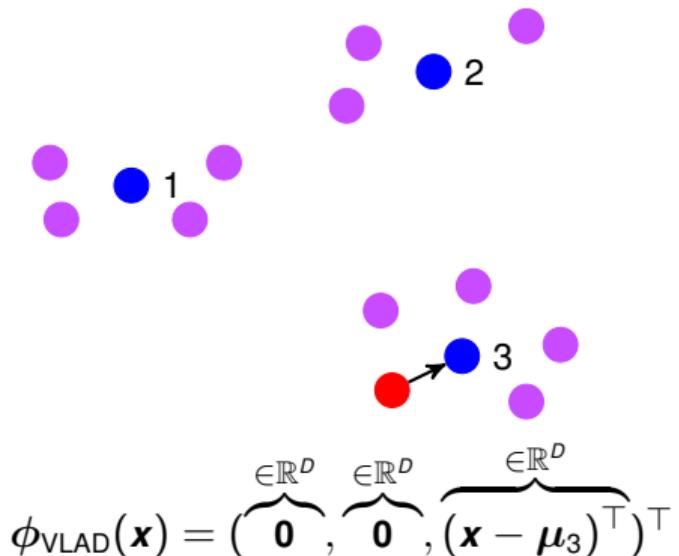


Local Descriptors:  $\mathcal{X} = \{\mathbf{x}_i \in \mathbb{R}^D, i = 1, \dots, T\}$   
Clusters:  $\mathcal{D} = \{\boldsymbol{\mu}_k \in \mathbb{R}^D, k = 1, \dots, K\}$

<sup>4</sup>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<sup>4</sup>



Local Descriptors:  $\mathcal{X} = \{\mathbf{x}_i \in \mathbb{R}^D, i = 1, \dots, T\}$   
 Clusters:  $\mathcal{D} = \{\boldsymbol{\mu}_k \in \mathbb{R}^D, k = 1, \dots, K\}$

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

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

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

<sup>4</sup>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.



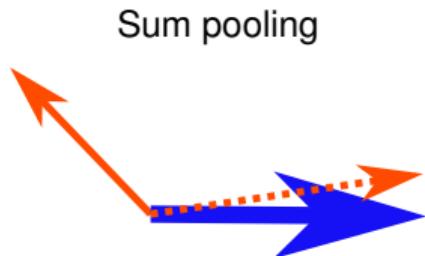
**FAU**

FRIEDRICH-ALEXANDER-  
UNIVERSITÄT  
ERLANGEN-NÜRNBERG  
SCHOOL OF ENGINEERING

# Sum Pooling vs. Generalized Max Pooling

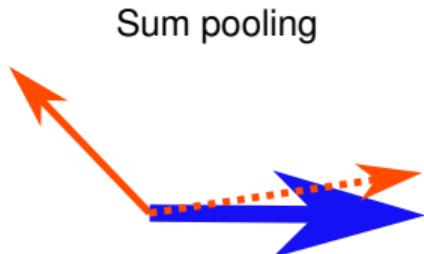


## Visual Burstiness



- Unrelated descriptors produce interference
- Frequent descriptors dominate similarity

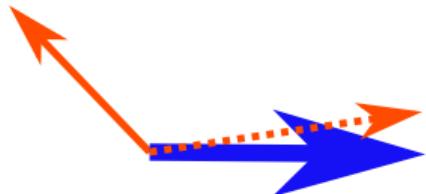
## Visual Burstiness



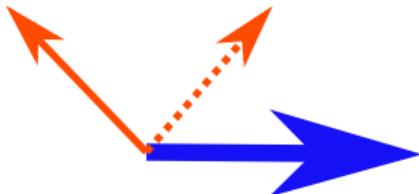
- Unrelated descriptors produce interference
- Frequent descriptors dominate similarity
- Choose better embedding
- Normalize encoding
  - Power normalization
  - Intra normalization
  - ...

## Visual Burstiness

Sum pooling



Generalized max pooling [5]



- Unrelated descriptors produce interference
- Frequent descriptors dominate similarity
- Choose better embedding
- Normalize encoding
  - Power normalization
  - Intra normalization
  - ...

→ Balance pooling

## Generalized Max Pooling

- Seek encoding  $\xi$  which weights each embedding  $\phi$

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

Generalized max pooling [5]



## Generalized Max Pooling

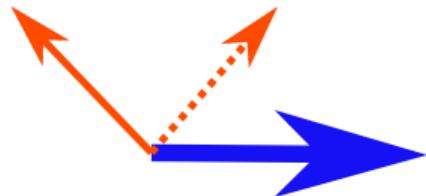
- Seek encoding  $\xi$  which weights each embedding  $\phi$

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

- Max pooling: equally similar to frequent and rare patches
- Enforce similarity between any patch encoding and aggregated representation to be constant

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

Generalized max pooling [5]



## Generalized Max Pooling

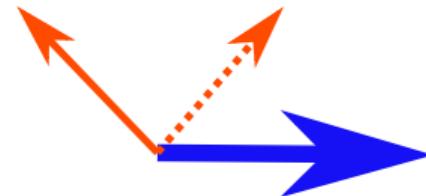
- Seek encoding  $\xi$  which weights each embedding  $\phi$

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

- Max pooling: equally similar to frequent and rare patches
- Enforce similarity between any patch encoding and aggregated representation to be constant

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

Generalized max pooling [5]



→ Optimization problem can be cast as a ridge regression problem

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

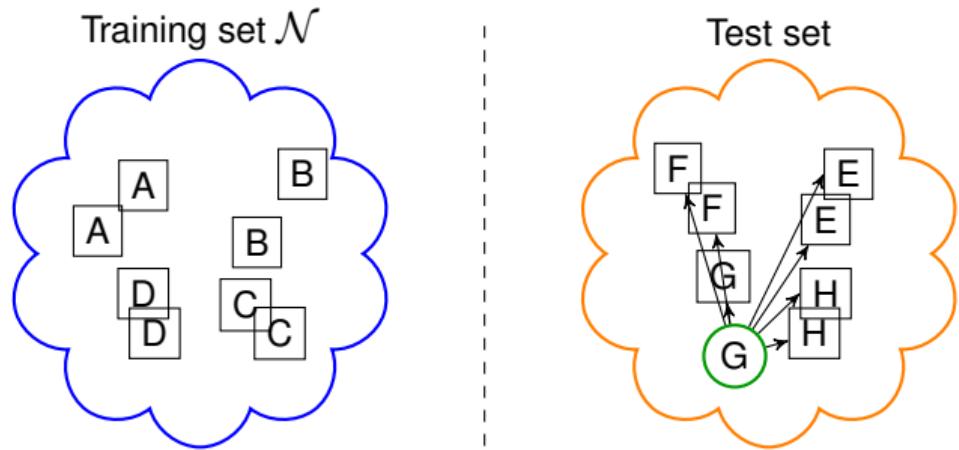
$\lambda \rightarrow 0$ : max pooling  
 $\lambda \rightarrow \infty$ : sum pooling



# Exemplar Classification

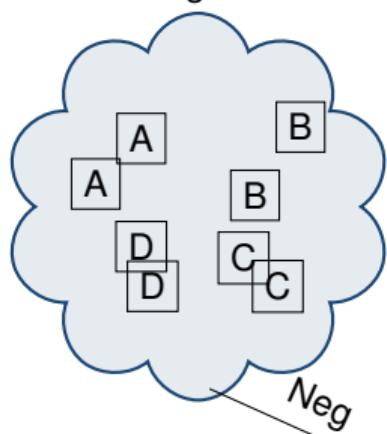


## Similarity

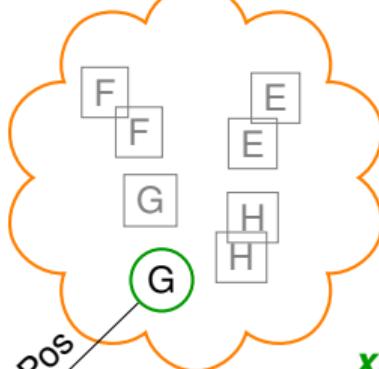


## Similarity

Training set  $\mathcal{N}$



Test set



### Exemplar SVMs [Christlein17a]

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

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

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

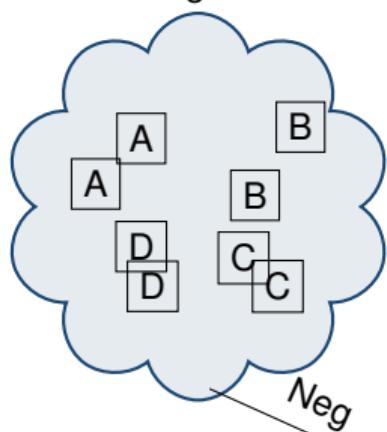
$c_p, c_n$  : margin parameters

(e.g. indirect proportional to #samples)

⇒ Subject-specific similarity!

## Similarity

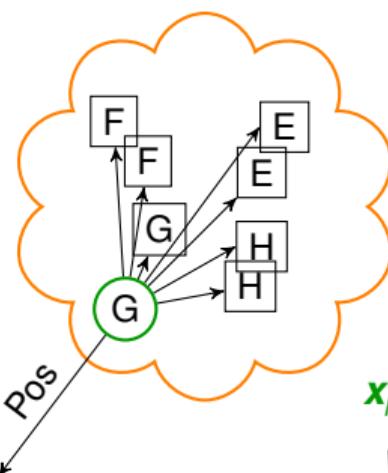
Training set  $\mathcal{N}$



E-SVM

⇒ Subject-specific similarity!

Test set



**Exemplar SVMs** [Christlein17a]

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

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

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

$c_p, c_n$  : margin parameters

(e.g. indirect proportional to #samples)

## Exemplar SVMs as Feature Encoder

- Trained SVM models:  $w$ ,  $b$
  - Similarity independent of  $b$  when using cosine distance
- ⇒ New feature:  $x \rightarrow \frac{w}{\|w\|_2}$
- Note: iterative application of E-SVM possible but benefit vanishes quickly

Thank you for your attention



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. Papandreu, "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.