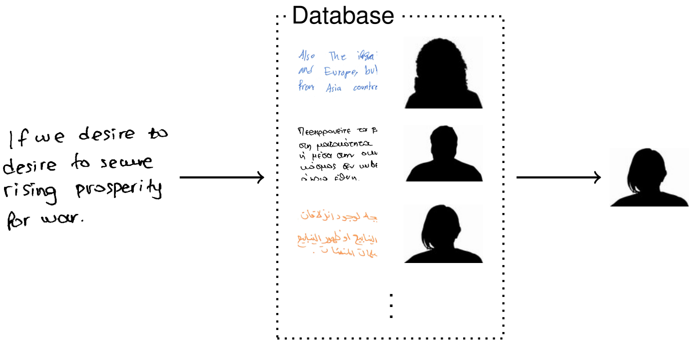
FRIEDRICH-ALEXANDER-
UNIVERSITÄT
ERLANGEN-NÜRNBERG
SCHOOL OF ENGINEERING

# Unsupervised Feature Learning for Writer Identification and Writer Retrieval

**Vincent Christlein**, Martin Gropp, Stefan Fiel, Andreas Maier
Pattern Recognition Lab, Friedrich-Alexander University of Erlangen-Nuremberg
November 15th, 2017

# Writer Identification vs. Writer Retrieval

# Writer Identification vs. Writer Retrieval




## Writer Identification

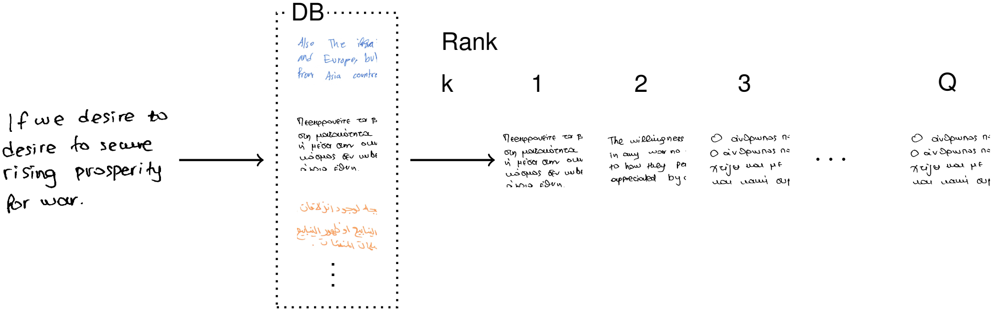
Given:

- Query document
- Documents of known writers

Wanted:

- Writer-ID

# Writer Identification vs. Writer Retrieval




## Writer Retrieval

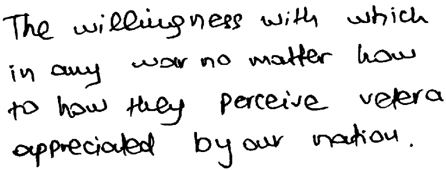
Given:

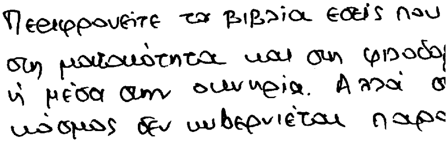
- Query document
- Documents of unknown writers

Wanted:

- Most similar documents

# Contemporary Datasets



## ICDAR13 benchmark dataset

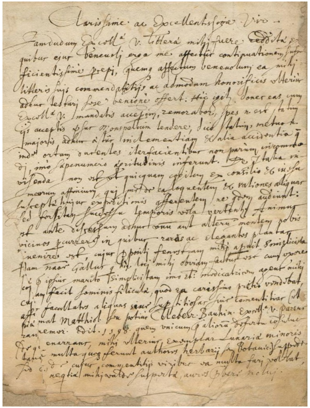
- 4 documents per writer (2 English, 2 Greek)
- Train: 100 writers
- Test: 250 writers

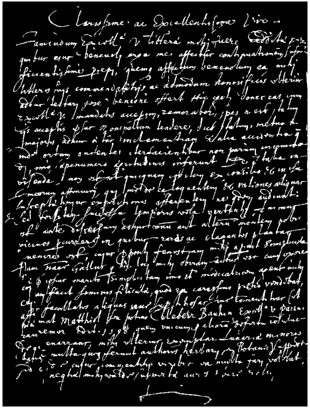
Other datasets: CVL (English, German), KHATT (Arabic), IAM (English)

# Historical Dataset

## ICDAR17 competition benchmark dataset

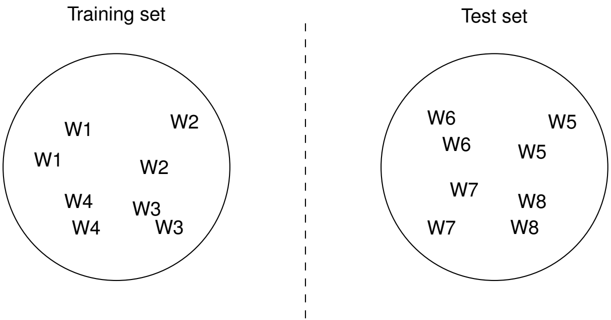
- Collection from university library Basel
- Train: 394 writers x 3 images
  $\rightarrow$ 1182 images
- Test: 720 writers x 5 images
  $\rightarrow$ 3600 images
- Additionally provided:
  - Manual text crops
  - Binarization (estimation)





S. Fiel, F. Kleber, M. Diem, *et al.*, "ICDAR 2017 Competition on Historical Document Writer Identification (Historical-WI)", , in *ICDAR*, 2017

# Writer-Independent Datasets

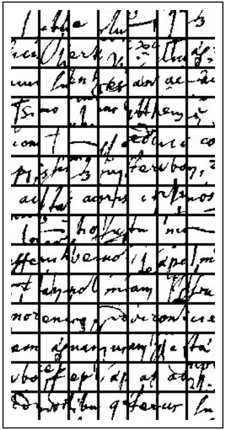


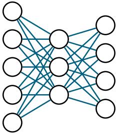

Training and test sets are independent

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

# Typical Methodology For Deep Learning Feature Extraction

# CNN Activation Features



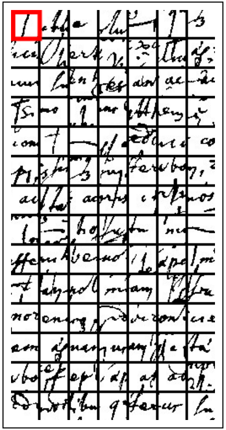


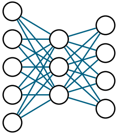
(Targets)

Surrogate classes:

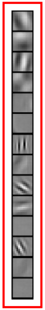
writers of train set

# CNN Activation Features

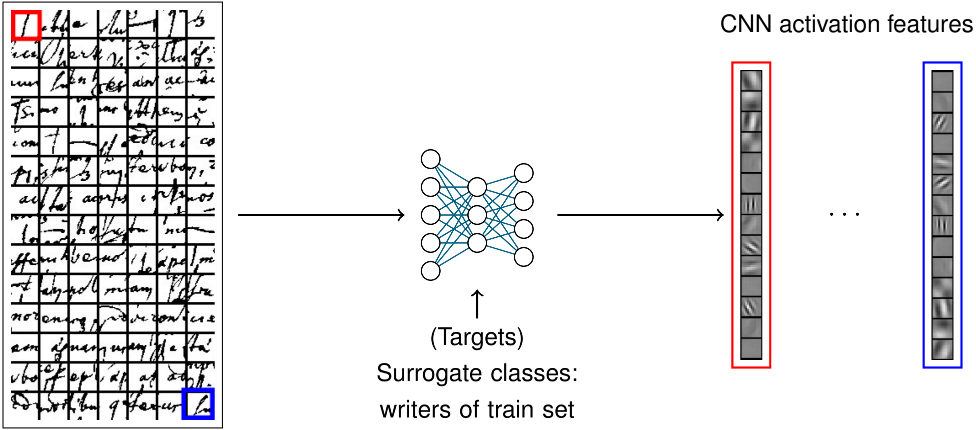




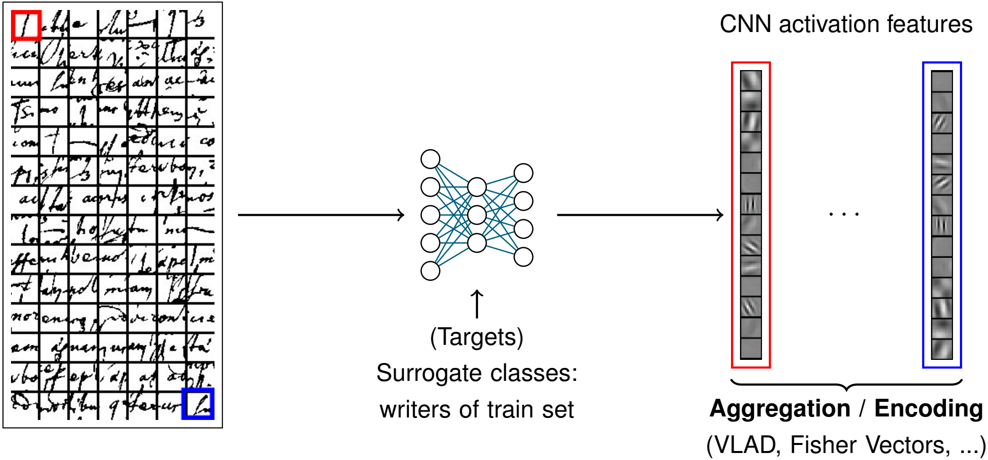
(Targets)
Surrogate classes:
writers of train set



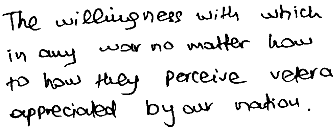
# CNN Activation Features

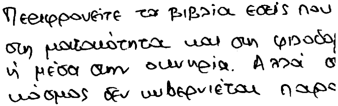


## CNN Activation Features



## Results for ICDAR13



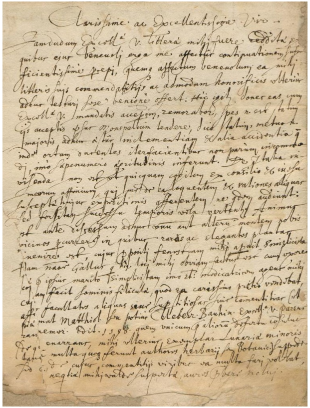




| Publication | Year | TOP-1 [%] | mAP [%] |
|---|---|---|---|
| Fiel & Sablatnig, CAIP | 2015 | 88.5 | – |
| Christlein et al., GCPR | 2015 | 98.9 | 88.6 |
| Tang & Wu, ICFHR | 2016 | 99.0 | – |

# Results for ICDAR17

## Train on official ICDAR17 training set

| | TOP-1 [%] | mAP [%] |
|---|---|---|
| Zernike[1] | 86.0 | 69.2 |

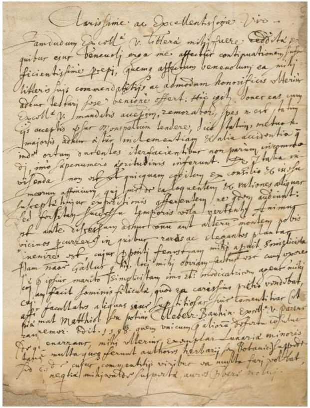


[1] V. Christlein, D. Bernecker, and E. Angelopoulou, “Writer Identification Using VLAD Encoded Contour-Zernike Moments”, in *ICDAR*, 2015

# Results for ICDAR17

## Train on official ICDAR17 training set

| | TOP-1 [%] | mAP [%] |
|---|---|---|
| Zernike[1] | 86.0 | 69.2 |
| CNN-AF (ResNet-18) | 67.4 | 46.1 |



[1] V. Christlein, D. Bernecker, and E. Angelopoulou, "Writer Identification Using VLAD Encoded Contour-Zernike Moments", in *ICDAR*, 2015

# Results for ICDAR17

## Train on official ICDAR17 training set

| | TOP-1 [%] | mAP [%] |
|---|---|---|
| Zernike[1] | 86.0 | 69.2 |
| CNN-AF (ResNet-18) | 67.4 | 46.1 |
| CNN-AF (LeNet) | 66.2 | 44.9 |

- Presumably: overfit on the training writers

→ Features don't generalize

→ Tried: other networks, models from earlier epochs, other encoding methods, ...

[1] V. Christlein, D. Bernecker, and E. Angelopoulou, "Writer Identification Using VLAD Encoded Contour-Zernike Moments", in *ICDAR*, 2015

# Results for ICDAR17

## Train on official ICDAR17 training set

| | TOP-1 [%] | mAP [%] |
|---|---|---|
| Zernike[1] | 86.0 | 69.2 |
| CNN-AF (ResNet-18) | 67.4 | 46.1 |
| CNN-AF (LeNet) | 66.2 | 44.9 |

## From the ICDAR17 competition results

| | | |
|---|---|---|
| Fribourg (ResNet-18) | 47.8 | 30.7 |

- Presumably: overfit on the training writers

→ Features don't generalize

→ Tried: other networks, models from earlier epochs, other encoding methods, ...

[1] V. Christlein, D. Bernecker, and E. Angelopoulou, "Writer Identification Using VLAD Encoded Contour-Zernike Moments", in *ICDAR*, 2015

# Results for ICDAR17

## Train on official ICDAR17 training set

| | TOP-1 [%] | $\mathrm{mAP}$ [%] |
|---|---|---|
| Zernike[1] | 86.0 | 69.2 |
| CNN-AF (ResNet-18) | 67.4 | 46.1 |
| CNN-AF (LeNet) | 66.2 | 44.9 |

## From the ICDAR17 competition results

| | | |
|---|---|---|
| Fribourg (ResNet-18) | 47.8 | 30.7 |

- Presumably: overfit on the training writers

$\rightarrow$ Features don't generalize

$\rightarrow$ Tried: other networks, models from earlier epochs, other encoding methods, ...

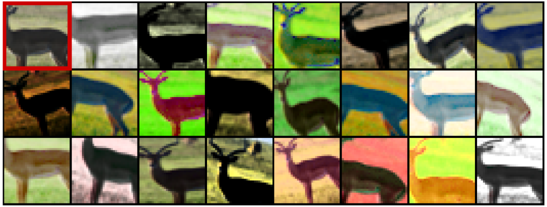
$\Rightarrow$ Not every surrogate task is useful

[1] V. Christlein, D. Bernecker, and E. Angelopoulou, "Writer Identification Using VLAD Encoded Contour-Zernike Moments", in *ICDAR*, 2015

# Unsupervised Feature Learning

# Related Work

Dosovitskiy et al. “Discriminative Unsupervised Feature Learning with Exemplar Convolutional Neural Networks”, PAMI, 2016

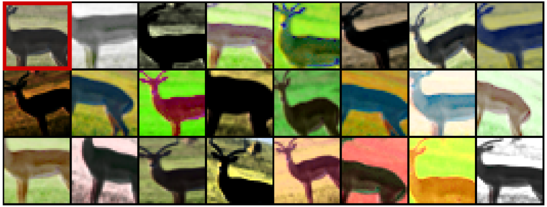


- Surrogate classes by random transformation

→ Each transformation = one class

# Related Work

Dosovitskiy et al. “Discriminative Unsupervised Feature Learning with Exemplar Convolutional Neural Networks”, PAMI, 2016

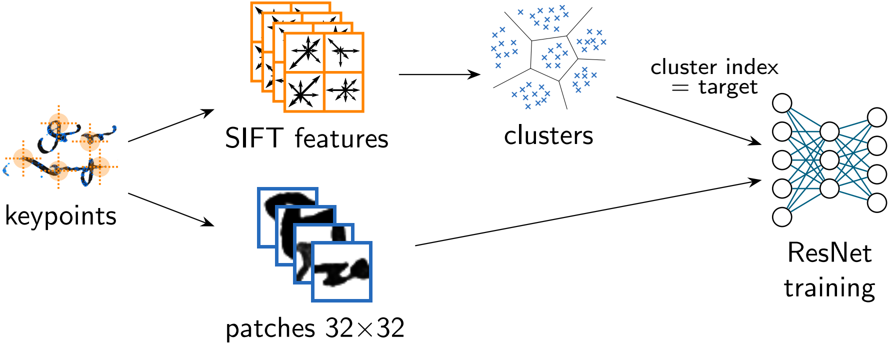


- Surrogate classes by random transformation

$\longrightarrow$ Each transformation = one class

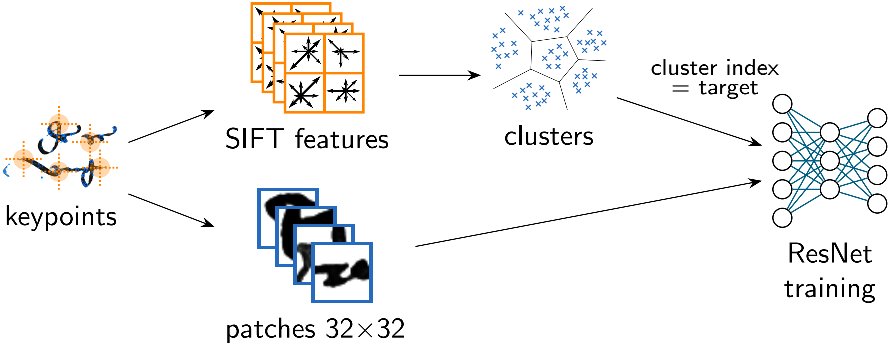
**But:** Script! $\longrightarrow$ Already have plenty of samples (= patches)!

# Unsupervised Feature Learning for Writer Recognition




- New surrogate classes: cluster indices of SIFT descriptors

# Unsupervised Feature Learning for Writer Recognition




- New surrogate classes: cluster indices of SIFT descriptors

→ Map similar patches to each other (kinda: autoencoder w.o. reconstruction)

\+ **No writer labels needed!**

# Performance

| | TOP-1 [%] | $\mathrm{mAP}$ [%] |
|---|---|---|
| Zernike | 86.0 | 69.2 |
| Cluster-CNN-AF (ResNet) | 87.3 | 72.4 |

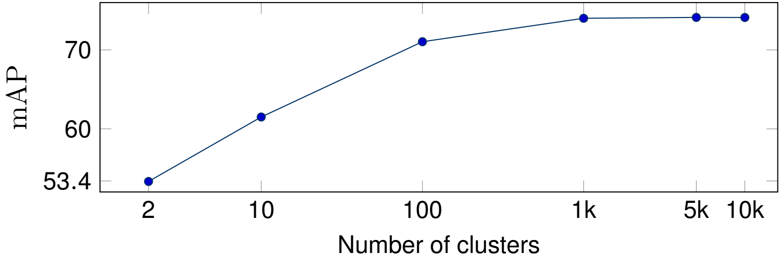
**Yeah!**

# Performance

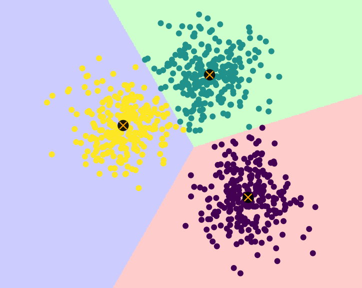
| | TOP-1 [%] | mAP [%] |
|---|---|---|
| Zernike | 86.0 | 69.2 |
| Cluster-CNN-AF (ResNet) | 87.3 | 72.4 |

**Yeah!**

## How many clusters / surrogate classes?



# Improve Baseline



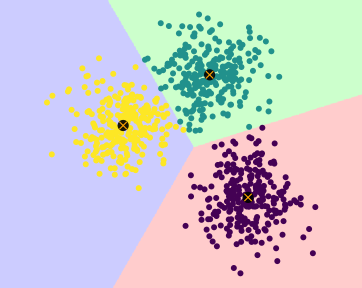
$\boldsymbol{\mu}_1$, $\boldsymbol{\mu}_2$: 1st, 2nd closest cluster centers

Distance ratio of descriptor **x**:

$$\rho = \frac{\|\mathbf{x} - \boldsymbol{\mu}_1\|_2}{\|\mathbf{x} - \boldsymbol{\mu}_2\|_2}$$

If $\rho > 0.9 \rightarrow$ remove **x** from training set

# Improve Baseline



$\boldsymbol{\mu}_1$, $\boldsymbol{\mu}_2$: 1st, 2nd closest cluster centers

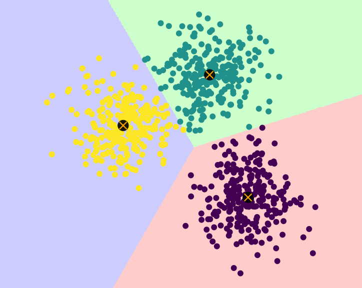
Distance ratio of descriptor $\mathbf{x}$:

$$\rho = \frac{\|\mathbf{x} - \boldsymbol{\mu}_1\|_2}{\|\mathbf{x} - \boldsymbol{\mu}_2\|_2}$$

If $\rho > 0.9 \rightarrow$ remove $\mathbf{x}$ from training set

| | TOP-1 [%] | mAP [%] |
|---|---|---|
| Cluster-CNN-AF | 87.3 | 72.4 |
| Cluster-CNN-AF ($\rho = 0.9$) | 88.3 | 74.1 |

# Improve Baseline



$\boldsymbol{\mu}_1, \boldsymbol{\mu}_2$: 1st, 2nd closest cluster centers

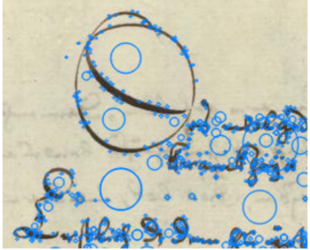
Distance ratio of descriptor **x**:

$$\rho = \frac{\|\mathbf{x} - \boldsymbol{\mu}_1\|_2}{\|\mathbf{x} - \boldsymbol{\mu}_2\|_2}$$

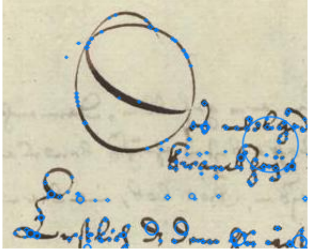
If $\rho > 0.9 \rightarrow$ remove **x** from training set

| | TOP-1 [%] | mAP [%] |
|---|---|---|
| Cluster-CNN-AF | 87.3 | 72.4 |
| Cluster-CNN-AF ($\rho = 0.9$) | 88.3 | 74.1 |
| Cluster-CNN-AF ($\rho = 0.9$, $L = 44$) | 88.2 | 74.3 |

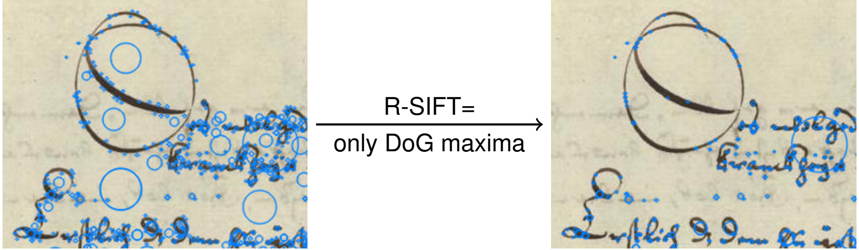
# SIFT Keypoints vs. Restricted SIFT Keypoints



R-SIFT= only DoG maxima

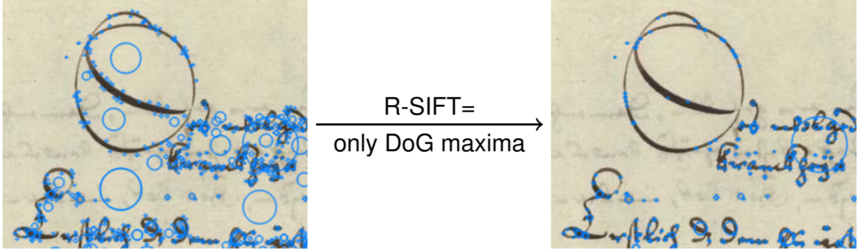


# SIFT Keypoints vs. Restricted SIFT Keypoints



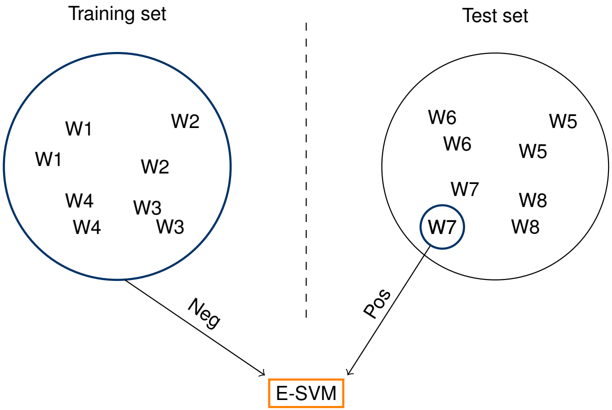

| | TOP-1 | mAP |
|---|---|---|
| Cluster-CNN-AF (Baseline: Binarized. / R-SIFT) | 88.3 | 74.1 |
| Cluster-CNN-AF (Binarized / SIFT) | 88.6 | 74.8 |

# SIFT Keypoints vs. Restricted SIFT Keypoints




| | TOP-1 | mAP |
|---|---|---|
| Cluster-CNN-AF (Baseline: Binarized. / R-SIFT) | 88.3 | 74.1 |
| Cluster-CNN-AF (Binarized / SIFT) | 88.6 | 74.8 |
| Cluster-CNN-AF (Grayscale / R-SIFT) | 87.1 | 71.6 |
| Cluster-CNN-AF (Grayscale / SIFT) | 87.7 | 72.3 |

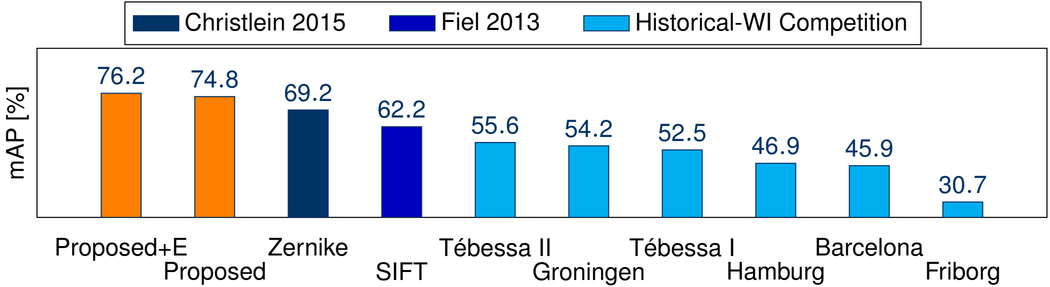
# Writer Adaptation Using Exemplar SVMs



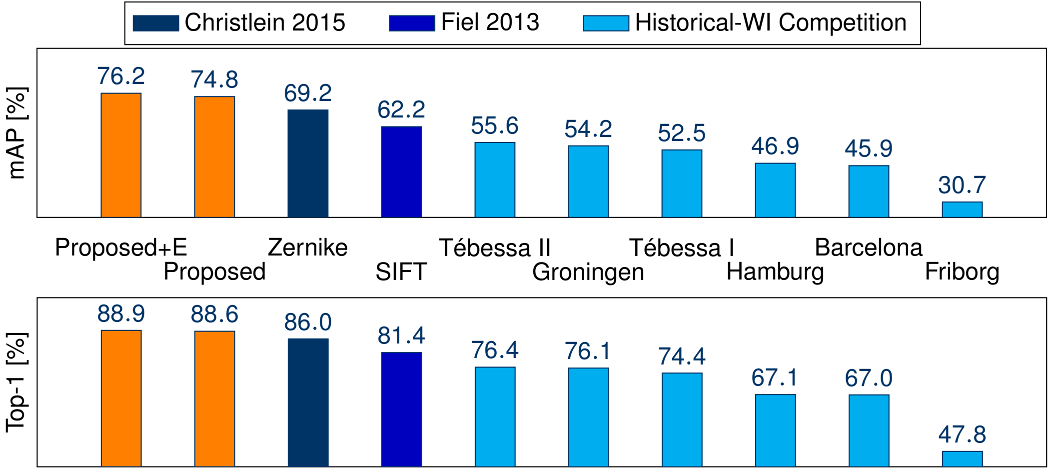

Compute individual E-SVMs $\Rightarrow$ individual similarity

V. Christlein, D. Bernecker, F. Hönig, *et al.*, "Writer Identification Using GMM Supervectors and Exemplar-SVMs", *Pattern Recognition*, vol. 63, 2017

# Comparison w. State of the Art

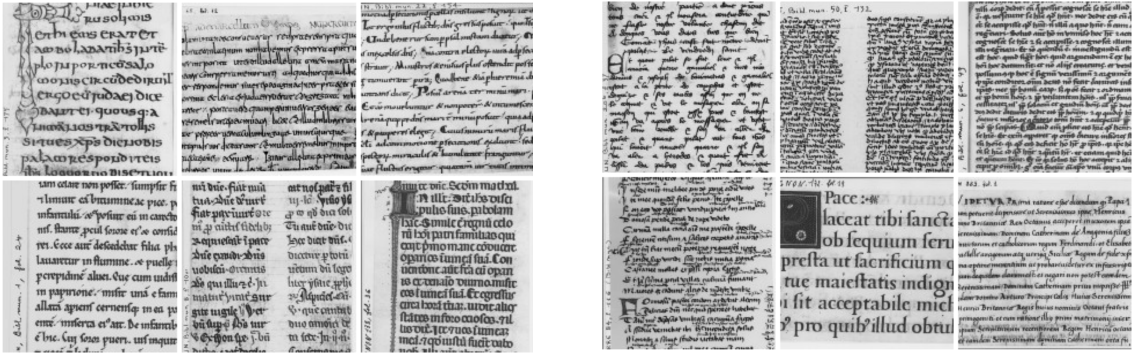


# Comparison w. State of the Art



# How does it generalize?

# Classification of Latin Medieval Manuscripts (CLaMM'16)



Source: http://clamm.irht.cnrs.fr/script-classes

- 2000 images, 12 script types (Uncial, Praegothica, Cursiva, . . .)

F. Cloppet, V. Églin, V. C. Kieu, *et al.*, "ICFHR2016 Competition on the Classification of Medieval Handwritings in Latin Script", in *ICFHR*, 2016, pp. 590–595

# Classification of Latin Medieval Manuscripts (CLaMM'16)

| Method | TOP-1 |
|---|---|
| DeepScript | 76.5 |
| FRDC-OCR | 79.8 |
| NNML | 83.8 |
| FAU | 83.9 |
| Cluster-CNN-AF + SVM | **84.1** |

Source: http://clamm.irht.cnrs.fr/script-classes

- 2000 images, 12 script types (Uncial, Praegothica, Cursiva, . . .)

F. Cloppet, V. Églin, V. C. Kieu, *et al.*, "ICFHR2016 Competition on the Classification of Medieval Handwritings in Latin Script", in *ICFHR*, 2016, pp. 590–595

# Classification of Latin Medieval Manuscripts (CLaMM'16)

| Method | TOP-1 |
|---|---|
| DeepScript | 76.5 |
| FRDC-OCR | 79.8 |
| NNML | 83.8 |
| FAU | 83.9 |
| Cluster-CNN-AF + SVM | **84.1** |
| Deeper CNN | 86.3 |

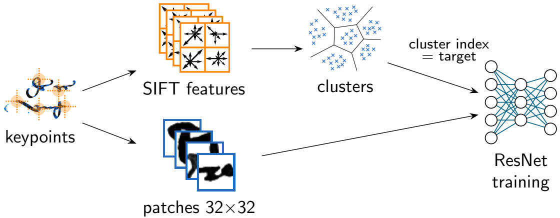
Source: http://clamm.irht.cnrs.fr/script-classes

- 2000 images, 12 script types (Uncial, Praegothica, Cursiva, . . .)

F. Cloppet, V. Églin, V. C. Kieu, *et al.*, "ICFHR2016 Competition on the Classification of Medieval Handwritings in Latin Script", in *ICFHR*, 2016, pp. 590–595

# Conclusion

# Conclusion




## Summary

- Proposed a new **unsupervised** feature learning technique for document analaysis
- Improves state of the art in writer identification / retrieval
- Generalizes well

## Outlook

- Try activations from other layers
- Incorporate text detection into the pipeline
- Add attention mechanism

https://github.com/vchristlein/icdar17code

**Questions?**