Recognizing Bengali Word Images - A Zero-Shot Learning Perspective

Sukalpa Chanda[1],Danïel Haitink[2], Jochem Baas[2], Prashant Kumar Prasad[3], Umapada Pal[3] and Lambert Schomaker[2]
[1]Østfold University College, Norway
[2]University of Groningen, The Netherlands
[3]Indian Statistical Institute, India

Motivation

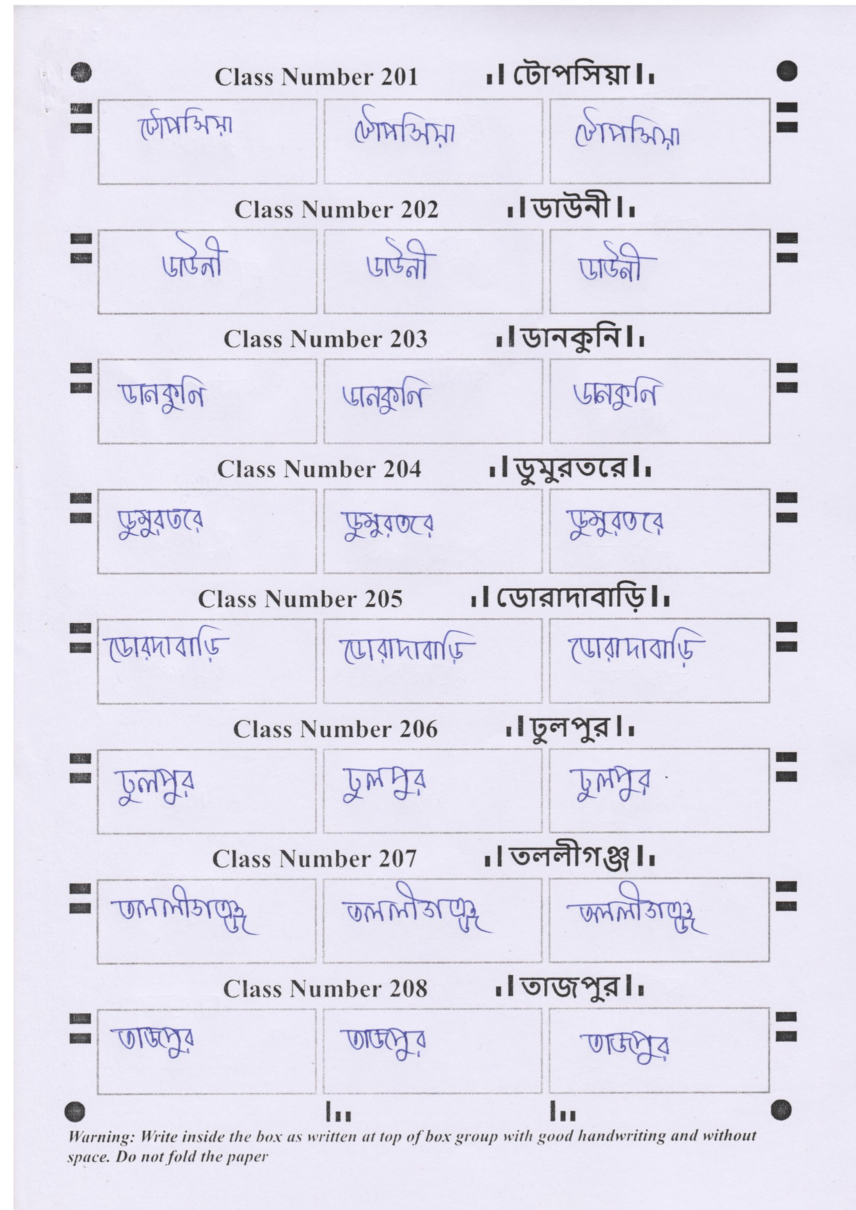
- Deep-learning-based methods are very popular and successful in different classification tasks
- But it demands labeled data for proper training
- Can only deal with “seen” class samples
- LSTMs can recognize “unseen” word classes, but requires fully transcribed text lines and sometimes a language model
- Labeling data demands human intervention, hence costly
- “Zero-shot” learning algorithms with proper feature and class attribute signature can counter this situation

Novelty/Challenges

- Zero-Shot Learning(ZSL) mainly has been explored for object detection
- To the best of our knowledge there is no work on any Indic script word recognition in ZSL perspective
- Signature/Semantic attribute space is very rich in object domain with information on colour and texture but such information is absent in handwritten text

Dataset

- 250 different word classes - those are place names in the State of West Bengal in India
- Data collection form contains 8 classes with space to provide 3 samples of handwriting for each class



Class Number 201 ।টোপসিয়া।

Class Number 202 ।ডাউনী।

Class Number 203 ।ডানকুনি।

Class Number 204 ।ডুমুরতরে।

Class Number 205 ।ডোরাদাবাড়ি।

Class Number 206 ।ঢুলপুর।

Class Number 207 ।তললীগঞ্জ।

Class Number 208 ।তাজপুর।

Warning: Write inside the box as written at top of box group with good handwriting and without space. Do not fold the paper

Dataset

- Elastic morphing based off-line data augmentation

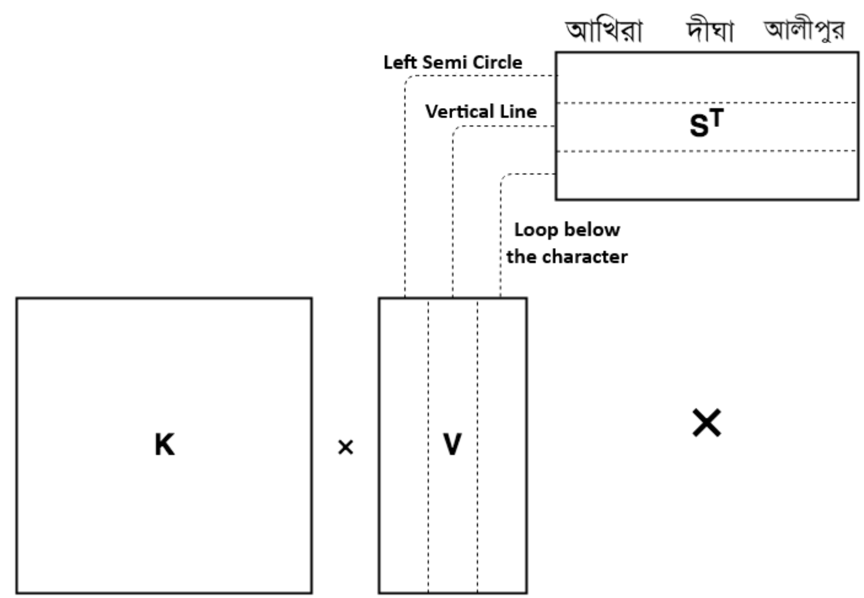
Data	Fold 0	Fold 1	Fold 2	Fold 3	Fold 4
Training	47360	47412	47300	47340	47370
Validation	11790	11800	11774	11780	11790
Testing	14796	14736	14868	14820	14787

Methodology

- Learning – is the mapping of basic shape attributes and deep features in matrix "V"
- K is a regular kernel matrix for example "Gaussian", "Polynomial" etc
- Classification - calculated per instance 'k' in K,where K could be a Gaussian Kernel or any other standard kernel function
- Classification - $_{\text{Argmax}}kVS_i^T$
- S_i is the signature attribute of i^{th} test class

জোড়াবাগান অনধিরামপাড়া দরিয়াপুর কালীঘাট

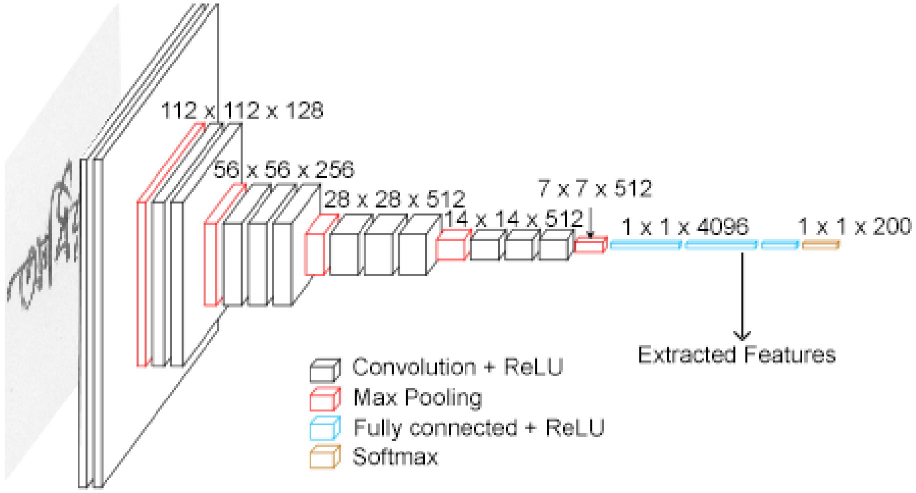
The basic shape attributes marked in red in different Bengali characters



Experimental Framework

- Five-fold cross validation with 50 test classes in each fold
- Different CNN architectures to generate features for word recognition

    - training from scratch
    - no data-flipping inside the architecture
- Features were extracted from output of FC1 layer of VGG16
- For InceptionNet, XceptionNet and ResNet, features were extracted from the average pool layer
- Deep-learned features along with shape attribute signature features are being fed to the Zero-shot learning algorithm



Schematic diagram of our customized VGG16 architecture as used in our experiment.

Results and Discussion

Performance with respect to different signature attributes

Sign. Attribute	Fold 0	Fold 1	Fold 2	Fold 3	Fold 4
S-Alph.	23.88%	32.35%	33.15%	29.66%	19.88%
4S-Sp.-Alph.	49.89%	39.06%	48.98%	49.06%	50.53%

Results and Discussion

Performance with respect to different CNN

Architecture	Fold 0	Fold 1	Fold 2	Fold 3	Fold 4
GoogleNet	35.09%	41.32%	30.28%	28.64%	39.66%
ResNet152	29.26%	28.52%	35.88%	26.07%	27.36%
XceptionNet	44.76%	35.45%	41.43%	38.21%	44.57%

Comparison

Method	Fold 0	Fold 1	Fold 2	Fold 3	Fold 4
AREN*	26.41%	27.24%	31.61%	25.11%	30.31%
Our Method	49.89%	39.06%	48.98%	49.06%	50.53%

* Guo-Sen Xie et al. “Attentive region embedding network for zero-shot learning,” in Proc. CVPR, 2019.

Conclusion

- “Unseen” word class images could be recognized using “Zero-shot” learning techniques with shape strokes as attribute signatures

- Efficacy of different CNN architectures were analyzed in the context of ZSL-based word image recognition

Questions!

- Please feel free to contact me during the poster session