Zero-Shot Image Classification via Hybrid Deep Learning and Semantic Embeddings

Shahariar Nafiz, C221361[1] and **Md. Meheraj Hossain, C221133**[1]

[1]Department of Computer Science and Engineering,, International Islamic University Chittagong, Bangladesh

Abstract

This paper presents a novel deep learning-based approach, DLIAP (Deep Learning-based Indirect Attribute Prediction), to improve zero-shot image classification by replacing traditional handcrafted features with deep representations. The method utilizes stacked sparse autoencoders (SSAE) and convolution operations for feature extraction. These features are integrated into an indirect attribute prediction (IAP) framework for classification. Experiments on Shoes, OSR, and a-Yahoo datasets demonstrate DLIAP's superior accuracy and generalization.

1 Introduction

Zero-shot learning (ZSL) addresses the classification of unseen object classes by leveraging shared semantic attributes. Traditional models such as DAP and IAP rely on low-level features (e.g., color, texture), which lack generalization across domains. The paper introduces deep learning for automatic, data-driven feature extraction to overcome these limitations, improving the knowledge transfer to unseen classes.

2 Methodology

2.1 Stage I: Network Training

- Image patches are extracted and ZCA-whitened to reduce pixel correlations.
- SSAE learns compressed representations, resulting in a feature mapping matrix.

2.2 Stage II: Feature Extraction

- The mapping matrix is used as a convolution kernel.
- Convolution and pooling reduce feature dimensionality and extract robust descriptors.

2.3 Stage III: Zero-Shot Classification

- Features are fed into the IAP model for attribute prediction.
- Classification is performed using maximum posterior estimates.

3 Current Model (DLIAP)

The model combines SSAE and CNN principles to form a deep, unsupervised feature extractor. Unlike standard CNNs, convolutional kernels are pre-trained using unsupervised SSAE, making it suitable for ZSL.

4 Results

DLIAP achieves higher classification accuracy and attribute prediction than traditional models. Table 1 shows a comparison across different models.

Table 1: Classification Accuracy (%) Comparison on Shoes Dataset

Method	**Accuracy (avg)**	**Time (h)**
IAP	21.91	1.04
Bow_IAP	21.90	0.08
SPM_IAP	22.38	0.05
DAP	29.90	1.39
DAN	31.20	3.19
DLIAP	**34.79**	0.76

5 Class Distribution

- **Shoes:** 10 classes, 14,658 images
- **OSR:** 8 classes, 2,688 images
- **a-Yahoo:** 12 classes, 2,644 images

6 Feature Space Visualization using PCA

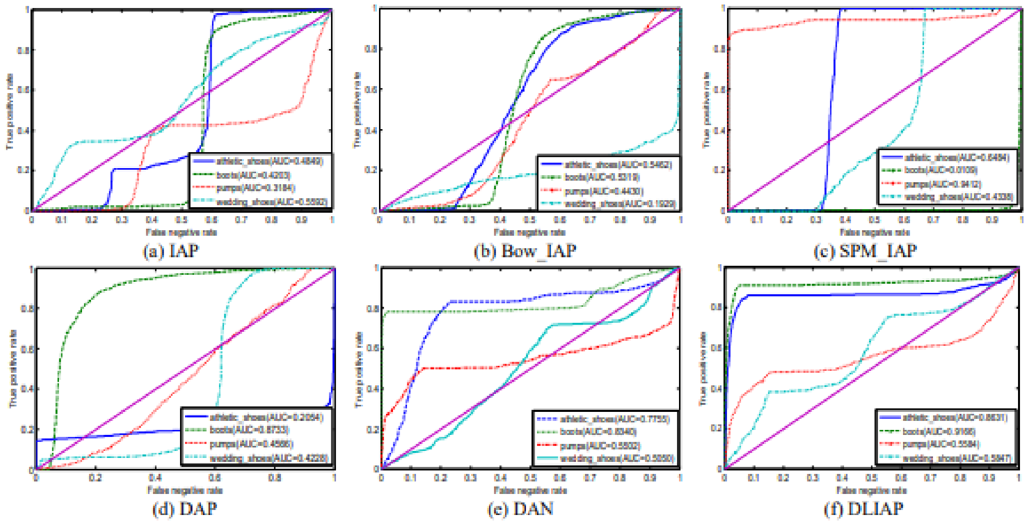
Although PCA is not explicitly visualized, dimensionality reduction is effectively achieved through pooling in CNN and SSAE compression, with optimal feature dimension being 576.

7 Model Performance

Performance improves with:

- Increased number of feature maps (best at 4)

- Increased image patches (optimal at 10,000)



(a) IAP (b) Bow_IAP (c) SPM_IAP

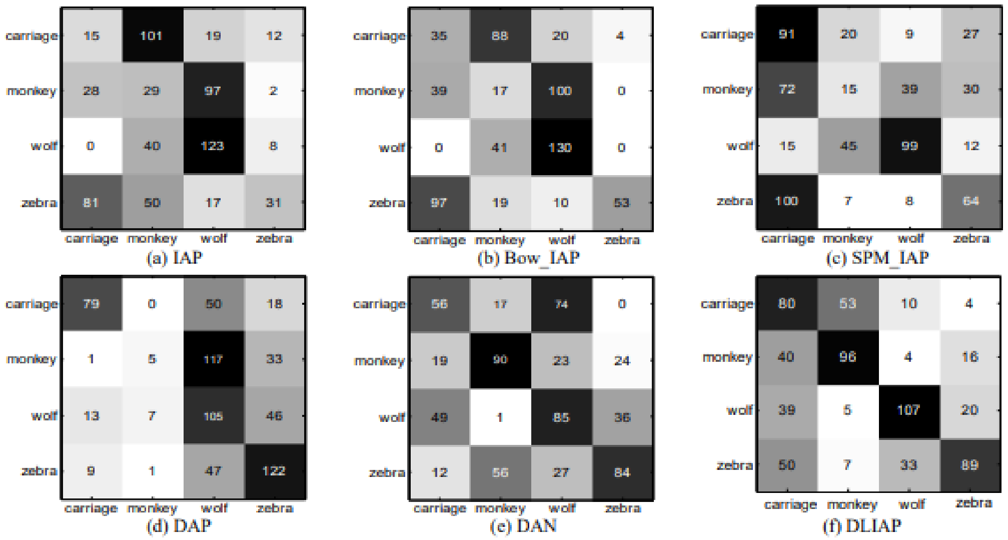
(d) DAP (e) DAN (f) DLIAP

Fig. 7 ROC curves and AUC values (Shoes dataset)

Figure 1: ROC curves and AUC values for different models on the Shoes dataset. Subplots: (a) IAP, (b) Bow_IAP, (c) SPM_IAP, (d) DAP, (e) DAN, (f) DLIAP.

8 Confusion Matrix

Confusion matrices demonstrate that DLIAP yields the fewest misclassifications. For instance, it accurately classifies most *athletic_shoes* images, unlike IAP and DAP, which often confuse visually similar categories.



	carriage	monkey	wolf	zebra
carriage	15	101	19	12
monkey	28	29	97	2
wolf	0	40	123	8
zebra	81	50	17	31

(a) IAP

	carriage	monkey	wolf	zebra
carriage	35	88	20	4
monkey	39	17	100	0
wolf	0	41	130	0
zebra	97	19	10	53

(b) Bow_IAP

	carriage	monkey	wolf	zebra
carriage	91	20	9	27
monkey	72	15	39	30
wolf	15	45	99	12
zebra	100	7	8	64

(c) SPM_IAP

	carriage	monkey	wolf	zebra
carriage	79	0	50	18
monkey	1	5	117	33
wolf	13	7	105	46
zebra	9	1	47	122

(d) DAP

	carriage	monkey	wolf	zebra
carriage	56	17	74	0
monkey	19	90	23	24
wolf	49	1	85	36
zebra	12	56	27	84

(e) DAN

	carriage	monkey	wolf	zebra
carriage	80	53	10	4
monkey	40	96	4	16
wolf	39	5	107	20
zebra	50	7	33	89

(f) DLIAP

Fig. 13 Confusion matrices of classification results (a-Yahoo dataset)

Figure 2: Confusion matrices of classification results on the Yahoo dataset. DLIAP shows superior performance with fewer misclassifications.

9 Comparative Observation

DLIAP achieves:

- +10% accuracy over IAP on the Shoes dataset
- Best ROC and AUC values across all test sets
- Higher robustness across domains

10 Discussion

DLIAP demonstrates that deep, unsupervised feature learning improves zero-shot classification by removing reliance on human-designed features. It generalizes across datasets, though at the cost of increased training complexity. Future work could explore optimizing efficiency and extending to multimodal ZSL settings.