Experimental Results of Autoencoder Features With/Without Fine-tuning

***Abstract*—**

***Keywords*-**

I. Introduction

In the previous experiments (ICDAR, HIP, and DAS), Kai, Oussama, and Me used Mathias' framework to extract autoencoder features. Generally, these features were learned by autoencoders without fine-tuning (if I am correct). In our HIP paper and also my thesis, we observed that: *a significant number of autoencoder features are redundant or irrelevant for layout classification.* This observation can be explained by the fact that autoencoders (without fine-tuning) are trained for reconstruction, rather than classification.

Autoencoders with fine-tuning are a kind of supervised feature learning technique (if I am correct). The learned features are supposed to be optimal or nearly optimal for classification. In other words, the learned features are supposed to be less redundant and less irrelevant for classification, compared to features learned by autoencoders without fine-tuning. This is an important motivation of our following experiments.

In the initial experiments, we investigate redundancy and irrelevance of features leaned by autoencoders with/without fine-tuning. We use autoencoder features to perform layout analysis, which classifies pixels on an image into five classes: out of page, background, comments, decorations, and text. Sequential Forward Selection (SFS) is used for feature selection. The results are shown in Table I.

Table I shows that for all classifiers using features learned by autoencoders with/without fine-tuning, generally feature dimension is decreased significantly, and comparable accuracy is obtained. So from Table I, we have two observations. (1) Regarding features learned with fine-tuning, a significant number of them are redundant or irrelevant for layout classification. This observation is similar with our previous observations about features learned by autoencoders without fine-tuning. (2) Regarding redundancy and irrelevance, there is no significant difference between features learned with and without fine-tuning. In other words, regardless of whether autoencoder features are learned with or without fine-tuning, they are equally redundant and irrelevant for layout classification in general.

Table I: Investigation of redundancy and irrelevance of features leaned by autoencoders with/without fine-tuning. The used classifiers are Support Vector Machine (SVM), Naive Bayes (NB), a decision algorithm (C4.5), and k-Nearest Neighbors (k-NN) where k=1.

Classifier	Autoencoders	Features	Dimension	Accuracy
SVM	With fine-tuning	Full features	46	53.57
		Selected features	24	53.99
	Without fine-tuning	Full features	46	54.99
		Selected features	23	55.28
NB	With fine-tuning	Full features	46	48.09
		Selected features	11	51.14
	Without fine-tuning	Full features	46	46.56
		Selected features	15	50.37
C4.5	With fine-tuning	Full features	46	53.64
		Selected features	14	53.82
	Without fine-tuning	Full features	46	55.61
		Selected features	13	54.50
1-NN	With fine-tuning	Full features	46	55.09
		Selected features	27	53.30
	Without fine-tuning	Full features	46	56.78
		Selected features	29	56.02