Learning of visualization of object recognition features and image reconstruction

Qiao Tan, Yuanlin Wen and Chenyue Meng

Background

We aim at visualizing and reconstructing images from Histogram of Oriented Gradient (HOG) features, i.e. recovering high-dimensional image vectors from low-dimensional feature vectors.

Introduction to HOG

HOG is a feature descriptor for object detection and recognition. It counts occurrences of gradient orientation in localized portions of an image.

Multi-class SVM

HOG features can be visualized by multi-class SVM algorithm with strategy known as 'One-Against-One'. This method constructs $\frac{k(k-1)}{2}$ classifiers where each one is trained on data from two classes. For training data from the i_{th} and the j_{th} classes, we solve the following binary classification problem:

$$
\begin{aligned}
\underset{\omega^{ij},b^{ij},\xi^{ij}}{\operatorname{argmin}} \quad & \frac{1}{2}(\omega^{ij})^T\omega^{ij} + C\sum_t \xi_t^{ij} \\
& (\omega^{ij})^T\phi(x_t) + b^{ij} \geq 1 - \xi_t^{ij}, \text{if} y_t = i, \\
& (\omega^{ij})^T\phi(x_t) + b^{ij} \leq -1 + \xi_t^{ij}, \text{if} y_t = j, \\
& \xi_t^{ij} \geq 0.
\end{aligned}
$$

We then use 'Max Wins' Strategy to choose between the label. As for the kernel, we find RBF performs best.

Pairwise Dictionary

To invert visual featrues and reconstruct the original images, we approximate an image $x_i = U\alpha$ and its HOG feature $\phi = V\alpha$ with a group of paired basis drawn from a overcomplete dictionary, which consists of paired natrual image basis $U \in \mathbb{R}^{P\times K}$ and feature space basis $V \in \mathbb{R}^{Q\times K}$. Once we obtain the paired dictionary from [2], we are able to use the matching pursuit algorithm [3] to estimate the shared coefficients α and estimate the original image based on the paired representation accordingly.

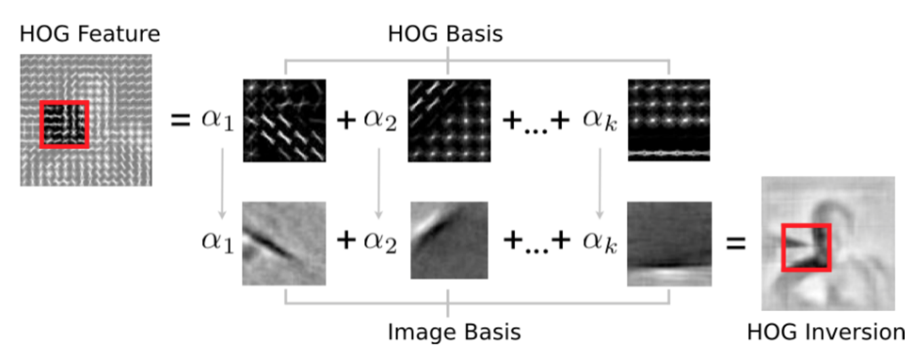

Figure 1: Inverting HOG feature using a paired dictionary.

Conv. Neural Network

I. HOG features fails to reconstruct images due to 1) position information is lost 2) computing HOG is different from CNN

II. CNN feature

Conv.NN: img $\rightarrow$ feature
Conv. + Maxpool + ReLU
deConv.NN: feature $\rightarrow$ img
Unpool + Conv. + ReLU [1]

Results

1. Multi-class SVM

Images shown in Figure 2 are visualizations of some images after training.

Figure 2: samples of HOG feature visualization by multi-class SVM algorithm. left: original images. right: predicted images

2. Pairwise Dictionary

We show some learned pair dictionary elements in Figure 3 and our inversion results in Figure 4 for a few object categories. Qualitatively, paired dictionary learning seems to produce very good visualization for HOG features. Moreover, previously unrevealed high frequency components in the original images are recovered through the inversion without introducing significant noise.

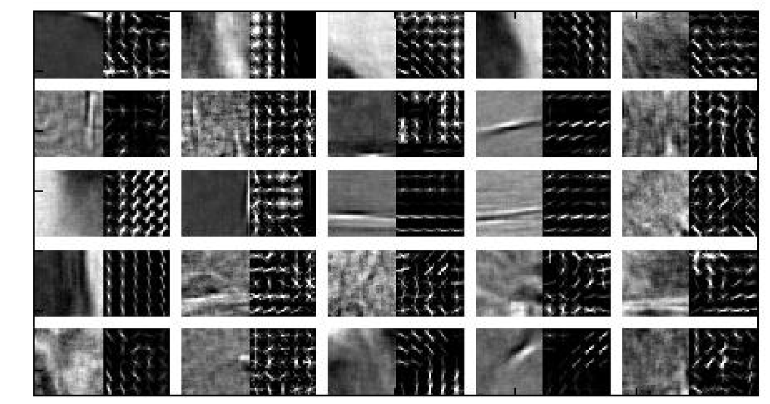

Figure 3: some of learned basis for feature and image spaces.

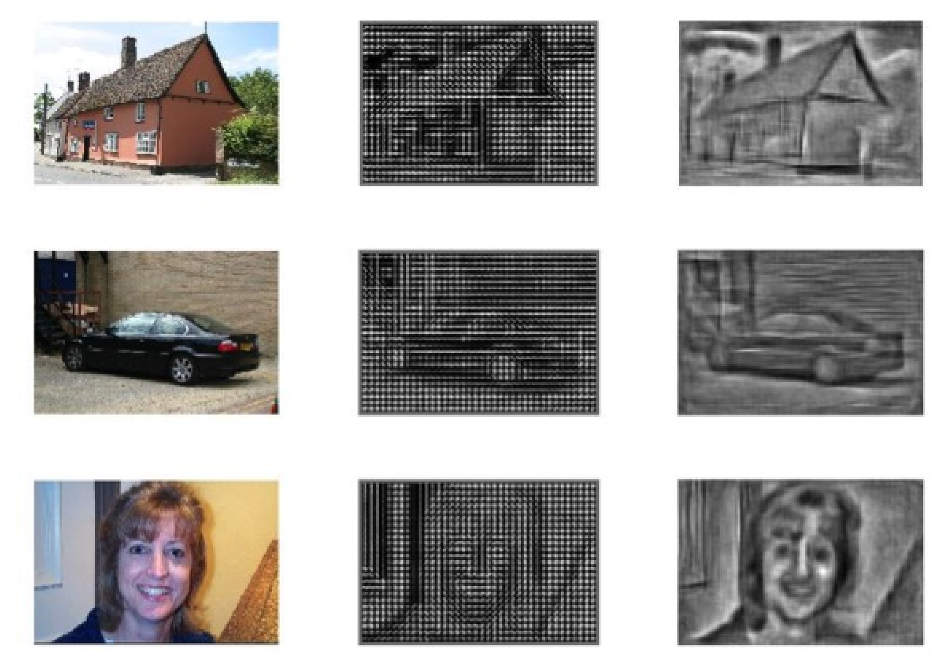

Figure 4: sample inversion results for different object recognition image databases. left: orginal images. middle: HOG features. right: inverted images.

3. CNN

Figure 5: CNN feature reconstruction results, from left to right: 1) original images, 2) reconstructed images (channel 8-4-8), 3) reconstructed images (channel 4-2-4)

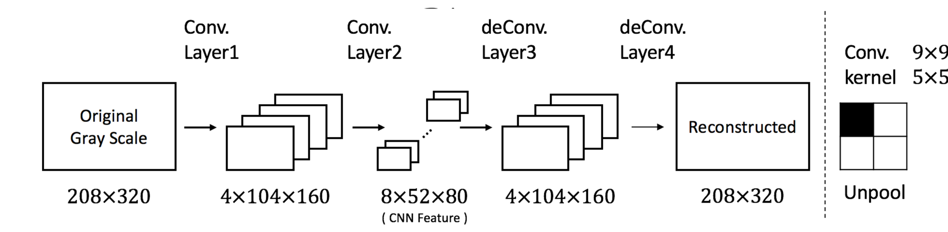

Figure 6: Architecture of convolutional and deconvolutional neural network

Figure 7: CNN feature visualization (of different number of channels), from top to bottom: 1) 8-4-8, 2) 4-4-4, 3) 4-2-4, 4) 8-2-8

Table 1: Mean normalized cross correlation

category	Multi-SVM	Pairwise Dict	CNN 8-4-8	CNN 4-2-4
house	0.933	0.593	0.835	0.430
face	0.925	0.636	0.756	0.261
car	0.895	0.639	0.846	0.169

References

[1] A. Dosovitskiy, J. T. Springenberg, and T. Brox. Learning to generate chairs with convolutional neural networks.

[2] J. Mairal, F. Bach, J. Ponce, and G. Sapiro. Online dictionary learning for sparse coding.

[3] S. G. Mallat and Z. Zhang. Matching pursuits with time-frequency dictionaries.