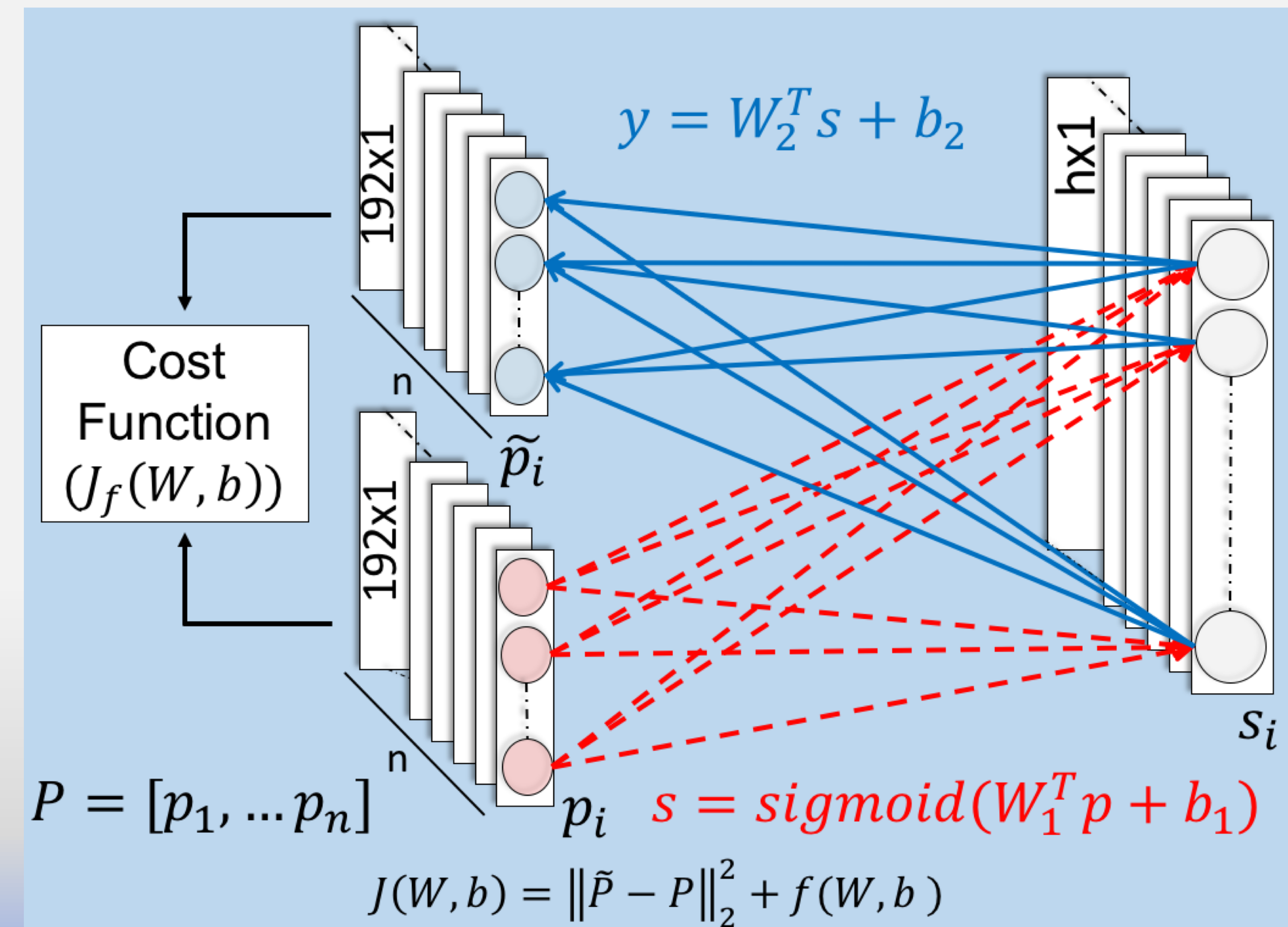
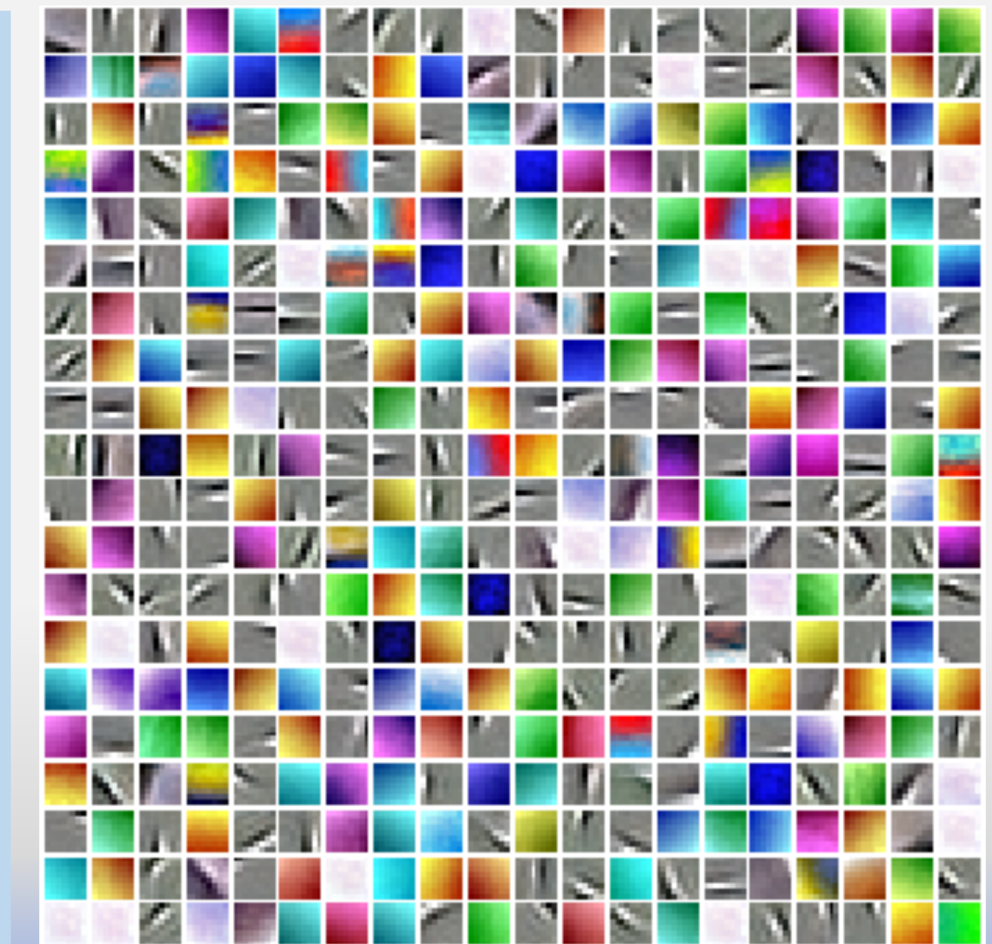
SEMANTICALLY INTERPRETABLE AND CONTROLLABLE FILTER SETS

Mohit Prabhushankar*, Gukyeong Kwon*, Dogancan Temel, and Ghassan AlRegib
Omni Lab for Intelligent Visual Engineering and Science (OLIVES)
School of Electrical and Computer Engineering , Georgia Institute of Technology
{mohit.p,gukyeong.kwon,cantemel,alregib}@gatech.edu

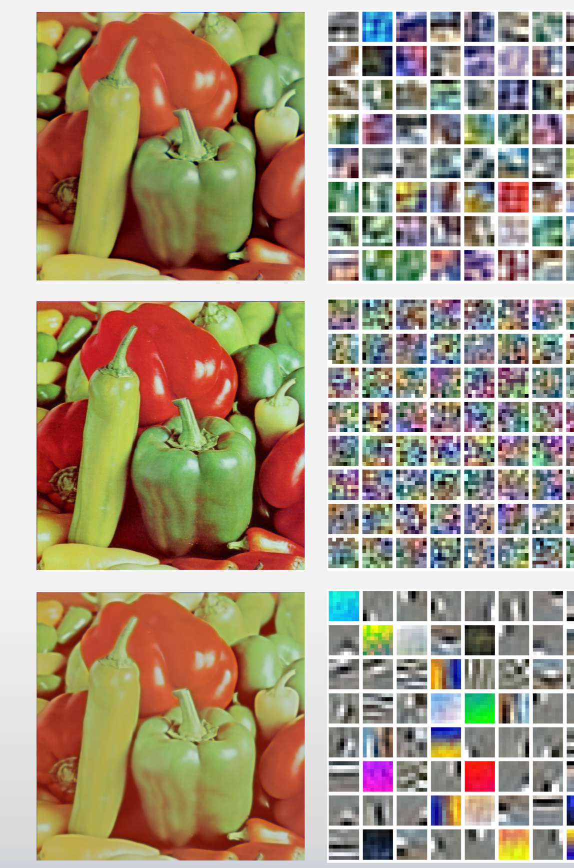
INTRODUCTION

- In this work, we generate and control semantically interpretable filters that are directly learned from natural images in an unsupervised fashion
- Each semantic filter learns a visually interpretable local structure in conjunction with other filters
- The significance of learning these interpretable filter sets is demonstrated on two contrasting applications
- The first application is image recognition under progressive decolorization, in which recognition algorithms should be color-insensitive to achieve a robust performance
- The second application is image quality assessment where objective methods should be sensitive to color degradation





CONTROL



L_1 Regularization

$\lambda\|W\|_2^2$

L_2 Regularization

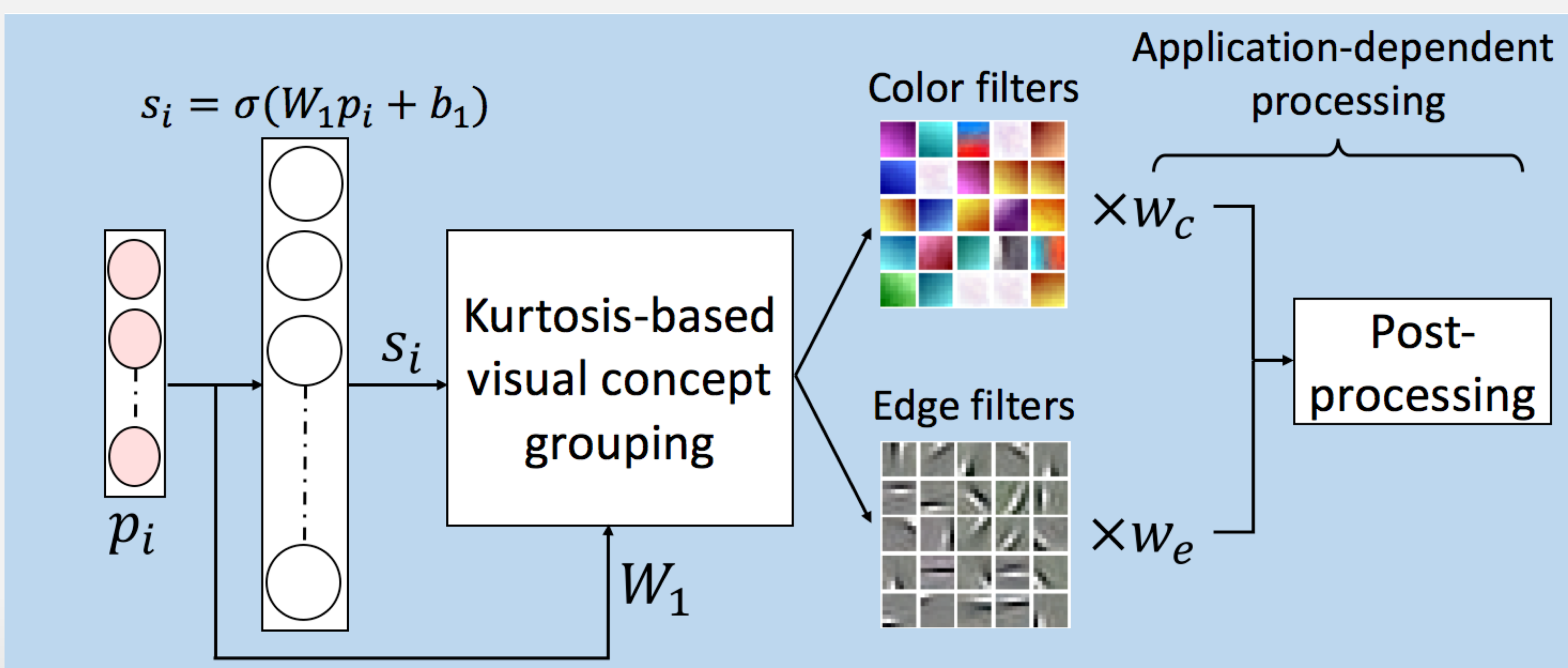
$\lambda\|W\|_2^2$

Elastic Net Regularization

$\beta\|W\|_1 + \lambda\|W\|_2^2$

Reconstruction quality $\xleftrightarrow{\text{Trade off}}$ Interpretability of filters

INTERPRET



WEIGHT GROUPING

$$\kappa[W_1] = E[W_1 - \mu]^4/\sigma^4$$
$$W_e = \kappa[W_1] > 0.5$$
$$W_c = \kappa[W_1] < 0.4$$

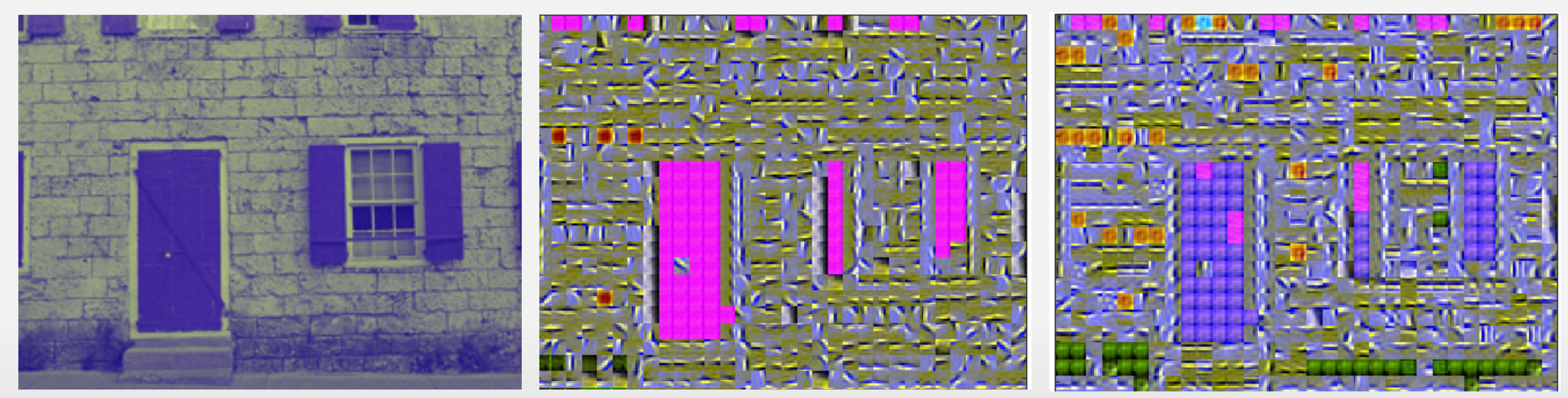
RESPONSE WEIGHING

$$s_e = w_e * [\sigma(W_e p_i + b_e)]$$
$$s_c = w_c * [\sigma(W_c p_i + b_c)]$$

APPLICATIONS

IMAGE QUALITY ASSESSMENT

Objective : To estimate the quality of a given distorted image, as perceived by humans

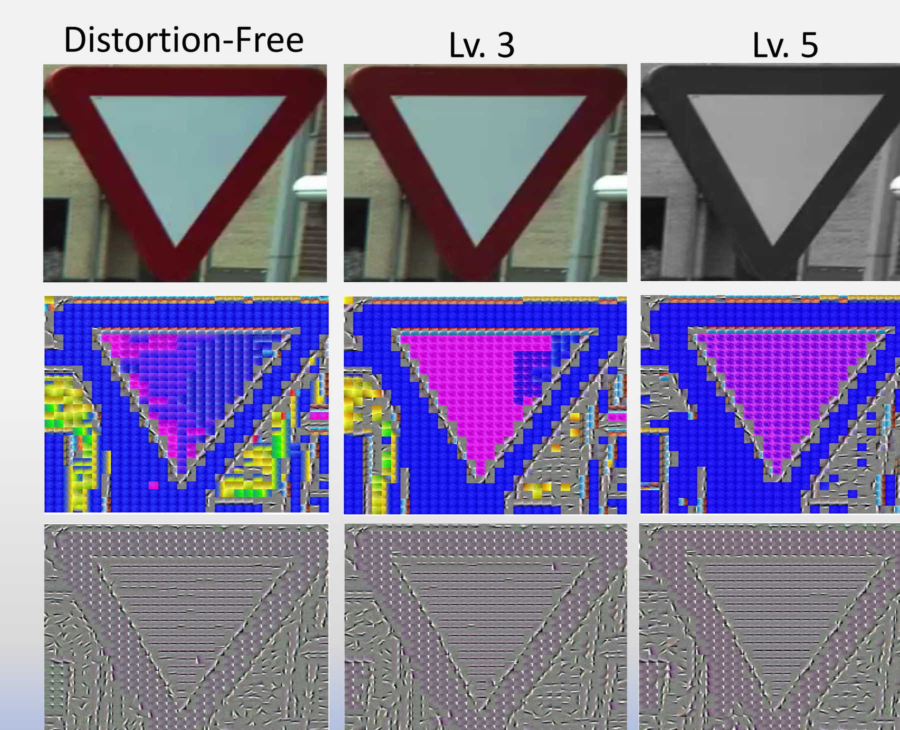


Original Image | Equal weights for edges and colors | Higher weights for edges compared to colors

- The performance is evaluated on Color Distortion challenges in TID 2013 database in terms of Pearson/Spearman correlation coefficients

TRAFFIC SIGN RECOGNITION

Objective : To recognize a give traffic sign under progressive decolorization.



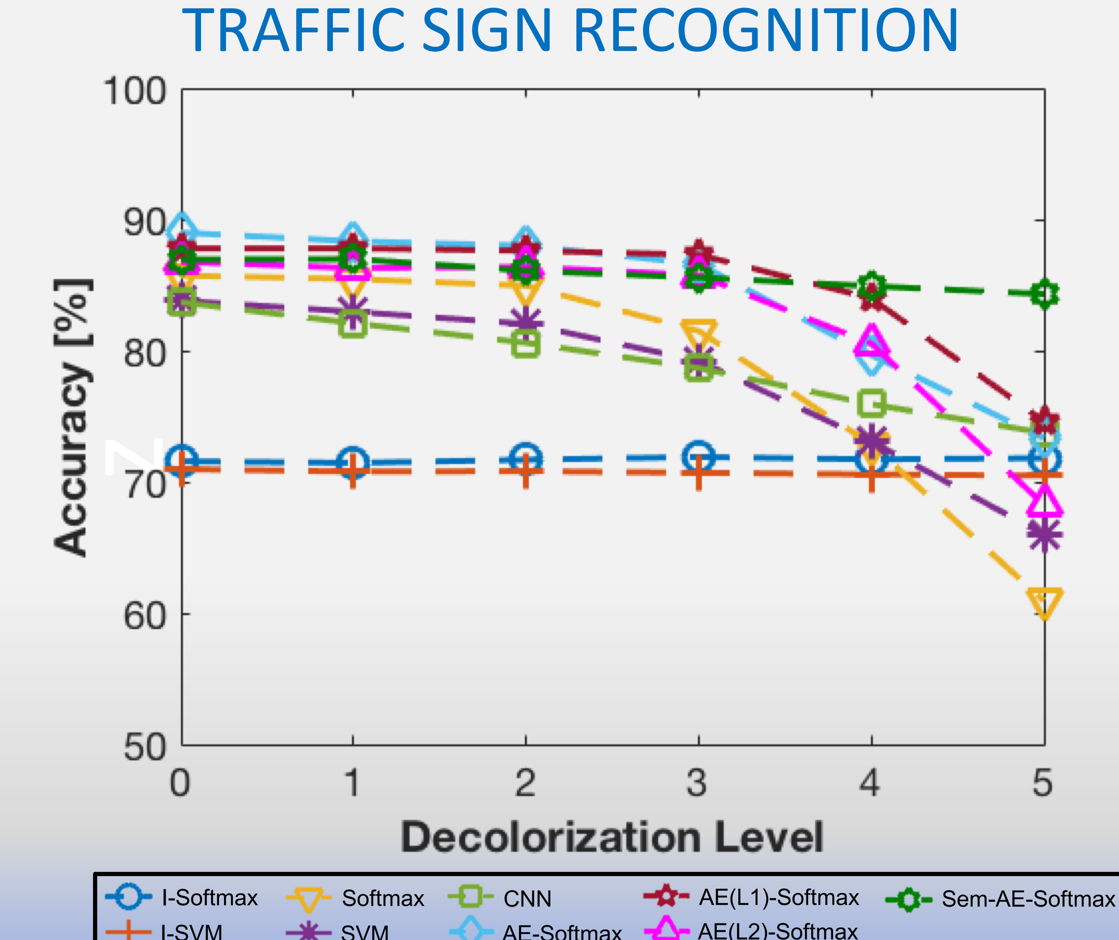
- CURE-TSR dataset which contains more than 2M real/unreal images with synthetic challenging conditions is utilized.
- We focus on decolorization challenge in this work.

RESULT

IMAGE QUALITY ASSESSMENT

Metric	Pearson Correlation Coefficient					Spearman Correlation Coefficient				
	Lv. 1	Lv. 2	Lv. 3	Lv. 4	Lv. 5	Lv. 1	Lv. 2	Lv. 3	Lv. 4	Lv. 5
PSNR-HMA	0.643	0.626	0.280	0.046	0.486	0.505	0.475	0.140	0.229	0.732
MS-SSIM	0.248	0.143	0.302	0.525	0.744	0.471	0.345	0.111	0.224	0.691
SR-SIM	0.370	0.260	0.301	0.497	0.732	0.505	0.401	0.098	0.234	0.732
FSIMc	0.391	0.253	0.303	0.553	0.778	0.432	0.347	0.013	0.395	0.793
PerSIM	0.126	0.085	0.304	0.554	**0.804**	0.306	0.160	0.143	0.479	**0.825**
AE	0.716	0.725	0.765	0.775	0.577	0.648	0.764	0.795	0.786	0.389
AE (L1)	0.557	0.406	0.542	0.682	0.619	0.451	0.378	0.541	0.660	0.480
AE (L2)	0.079	0.004	0.275	0.454	0.568	0.084	0.120	0.188	0.381	0.543
Sem-AE	**0.772**	**0.795**	**0.801**	**0.816**	0.730	**0.725**	**0.815**	**0.802**	**0.797**	0.615

TRAFFIC SIGN RECOGNITION



CONTURIBUTION

- We analyze various methods to control the training phase of an autoencoder. The filters learned from considered methods are visualized and validated based on their structural interpretability
- We group interpretable filter sets into semantically meaningful visual concepts that are based on color and edge characteristics
- We demonstrate the feasibility of semantic filter sets on two contrasting applications including image recognition and image quality assessment. Specifically, we test the robustness of these filters under mild to severe color degradation