

Introduction

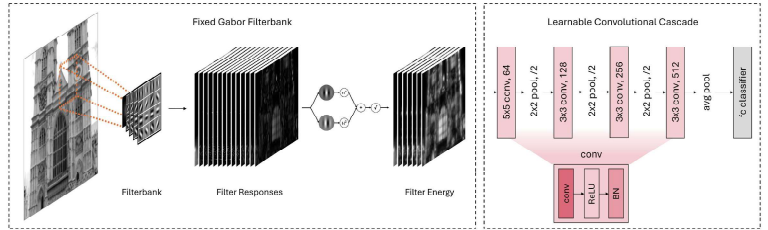
Scene gist: rapid, holistic understanding of a scene

Extracted in a single glance (< 100 ms)

Existing models¹ pool low-level image statistics

Gist extraction can improve with experience²

We propose ConvGist to model this improvement



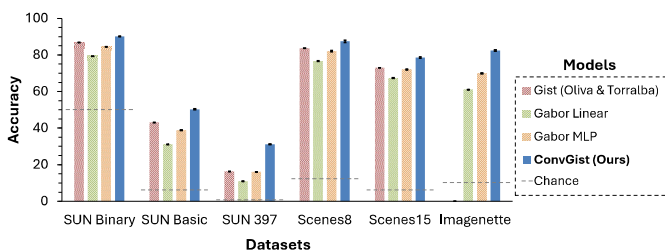
Proposed ConvGist Model: Fixed Gabor filterbank extracts low-level statistics, learnable cascade refines them for target domain

A Gabor filterbank paired with a compact convolutional cascade outperforms static gist models and much larger CNNs when data is limited.

ConvGist improves static gist representation

All models trained on each dataset separately

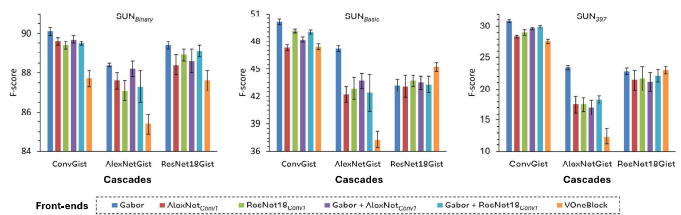
Gap widens on fine-grained tasks (e.g. SUN397³)



Gabor filterbank outperforms alternative frontends

Tested against Gabor-based VOneBlock⁵, and first layers of ImageNet-trained AlexNet and ResNet18

ConvGist pairing achieved highest scores

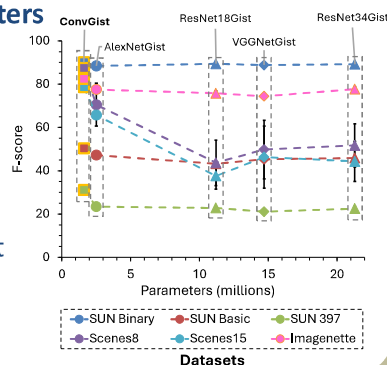


ConvGist cascade outperforms larger CNNs with a fraction of the parameters

4-layer cascade vs. AlexNet, VGG, ResNet

1.5 – 20x fewer parameters

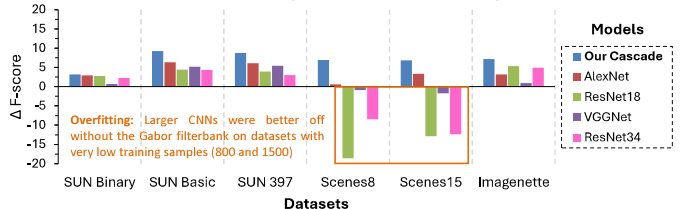
Larger backbones overfit on smaller datasets (Scenes8¹, Scenes15⁴)



Gabor filterbank is an effective inductive bias for data-efficient training

Standard CNNs vs same architectures with Gabor filterbank instead of first convolutional layer

Δ F-score = Model w/ Gabor – Model w/out Gabor



Future Work

Systematically evaluate component choices including cascade depth

Extend Gabor filterbank with simple cell responses and divisive normalization

Evaluate on large-scale datasets to test whether inductive bias persists

Compare model behavior against human gist perception

Evaluate on medical data to compare against medical experts

References

- Oliva & Torralba (2001). Int. J. Comput. Vis.
- Evans et al. (2013). Psychon. Bull. Rev.
- Xiao et al. (2016). Int. J. Comput. Vis.
- Lazebnik et al. (2006). CVPR.
- Dapello et al. (2020). NeurIPS.



Code



Profile