Object-Part Learning with Local Capsule Hierarchies

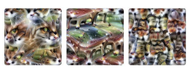
Nitish Dashora*[1], Silas Alberti*[1], Domas Buracas*[1], Bruno Olshausen[1]

[1]University of California, Berkeley

REDWOOD CENTER for Theoretical Neuroscience

Motivation

- Natural scenes have intrinsic structures like repetition of parts and spatial relationships.
- CNNs don't explicitly model this and are thus **uninterpretable** and **non viewpoint-invariant**.
- Capsules [2, 4] seek to model this structure in scenes by composing objects out of progressively more meaningful parts.
- We propose Local Capsule Hierarchies (LCH): an unsupervised generative model based on **hierarchical sparse coding** [3] which is inspired by the visual cortex of the brain.



Method – Single layer

A single-layer decomposes an image into a **sparse linear combination** σ of *N* **parts** with non-negative amplitudes σ.

Each part has its own **deformation parameter** represented by a unit vector θ. We model the deformation by decomposing the part into a linear combination in θ of *G* **learned basis functions** Φ.

$$I = \sum_{i=1}^{N} \sigma_i \sum_{j=1}^{G} \theta_{ij} \Phi_{ij} + \varepsilon$$

The layer outputs σ and θ are determined by minimizing a cost function comprised of reconstruction error (I) and a sparsity penalty for σ (II), while enforcing σ to be non-negative and letting θ only vary in its angle.

$$(\hat{\sigma}, \hat{\theta}) = \underset{\sigma \geq 0, \|\theta_i\|_2 = 1}{\arg\min} \underbrace{\left\| I - \sum_{i=1}^{N} \sigma_i \sum_{j=1}^{G} \theta_{ij} \Phi_{ij} \right\|_2}_{\text{(I)}} + \lambda \underbrace{\sum_{i=1}^{N} \sigma_i}_{\text{(II)}}$$

Method – Optimization

In practice, simply optimizing this cost function using **gradient descent** or ISTA/FISTA is **unstable**.

A solution is to use the **Subspace Locally Competitive Algorithm [1]** which is inspired by neuroscience. We first define the latent variable u:

$$u_{ij} = \sigma_i \theta_{ij} + \lambda \theta_{ij}$$

Then we optimize u using the following **dynamical system** which introduces **competition** between the parts through the inhibition term (III):

$$\tau \frac{\mathrm{d}u_{ij}}{\mathrm{d}t} = \langle I, \Phi_{ij} \rangle - \underbrace{\sum_{nm \neq ij} \sigma_n \theta_{nm} \langle \Phi_{ij}, \Phi_{nm} \rangle}_{\text{(III)}} - u_{ij}$$

The **basis functions** are learned through gradient descent on the reconstruction error while the inner optimization is unrolled with a small fixed number of iterations.

Results

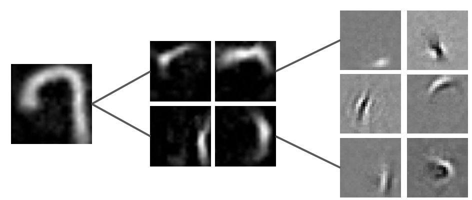


Figure 1: An example of LCH decomposition on the digit "7" from the MNIST dataset. The digit decomposes into mid-level parts which decompose into low-level parts

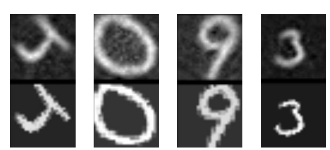


Figure 2: The reconstruction (top) versus the target (bottom) for the LCH training.

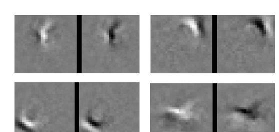
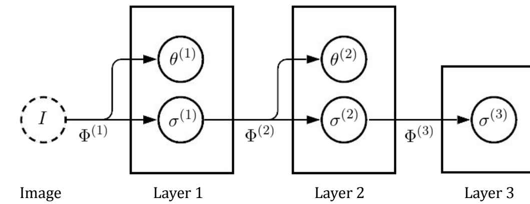


Figure 3: 4 LCH group examples with group size 2

Method – Multilayer

The higher layers decompose only the amplitudes σ, which can be interpreted as a local pooling operation over the learned deformation θ.



Future Work

- We seek to test this algorithm with downstream tasks such as classification or object detection. Given parse trees of scenes, LCH may prove to have certain strengths.
- Extending this to other datasets will test the robustness of LCH. Furthermore, adversarial analysis of LCH will expose potential strengths.
- Testing data efficiency of this method and experimenting with higher-level basis steering can show new applications.

Citations

[1]: Dylan M. Paiton, Steven Shepard, Kwan Ho Ryan Chan, and Bruno A. Olshausen. 2020. Subspace Locally Competitive Algorithms. In *Proceedings of the Neuro-inspired Computational Elements Workshop* (*NICE '20*). Association for Computing Machinery, New York, NY, USA, Article 9, 1–8.

[2]: Sara Sabour, Nicholas Frosst, and Geoffrey E. Hinton. 2017. Dynamic routing between capsules. In *Proceedings of the 31st International Conference on Neural Information Processing Systems* (*NIPS'17*). Curran Associates Inc., Red Hook, NY, USA, 3859–3869.

[3]: Olshausen BA, Field DJ. 1997. Sparse Coding with an Overcomplete Basis Set: A Strategy Employed by V1? *Vision Research, 37*: 3311-3325.

[4]: Adam R. Kosiorek, Sara Sabour, Yee Whye Teh, and Geoffrey E. Hinton. 2019. Stacked capsule autoencoders. *Proceedings of the 33rd International Conference on Neural Information Processing Systems.* Curran Associates Inc., Red Hook, NY, USA, Article 1390, 15512–15522.