Neural Network Pruning Denoises the Features and Makes Local Connectivity Emerge in Visual Tasks

F. Pellegrini, G. Biroli

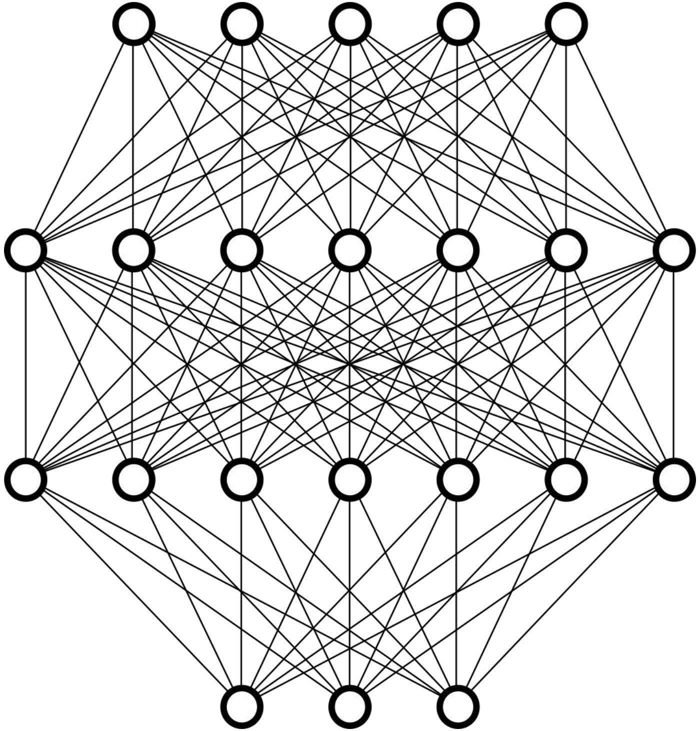
ENS | PSL

SIMONS FOUNDATION

PR[AI]RIE
PaRis Artificial Intelligence Research InstitutE

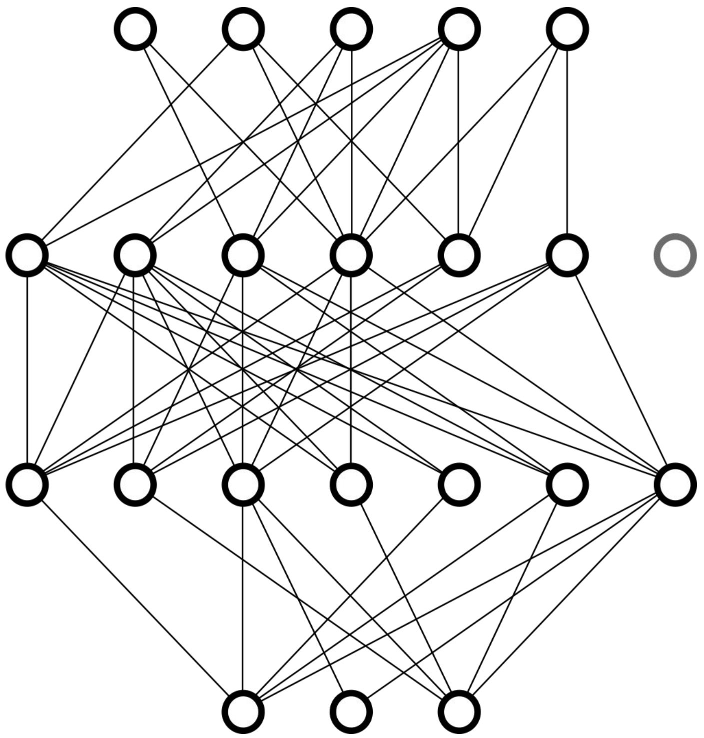
Network pruning

- Can some connections be removed?



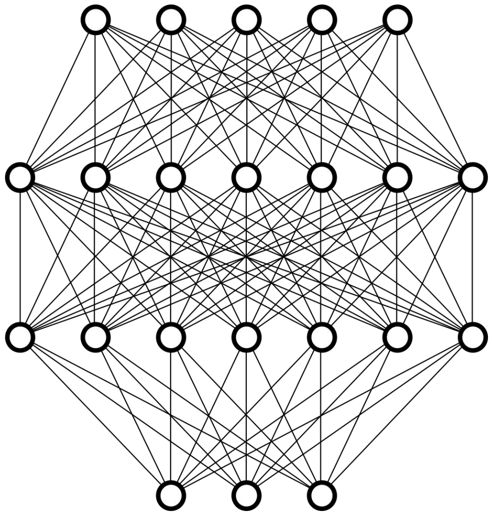
Network pruning

- Can some connections be removed?
- Possibly sparse result
- Multiple pruning criteria and schedules (curvature, magnitude, iterative…)



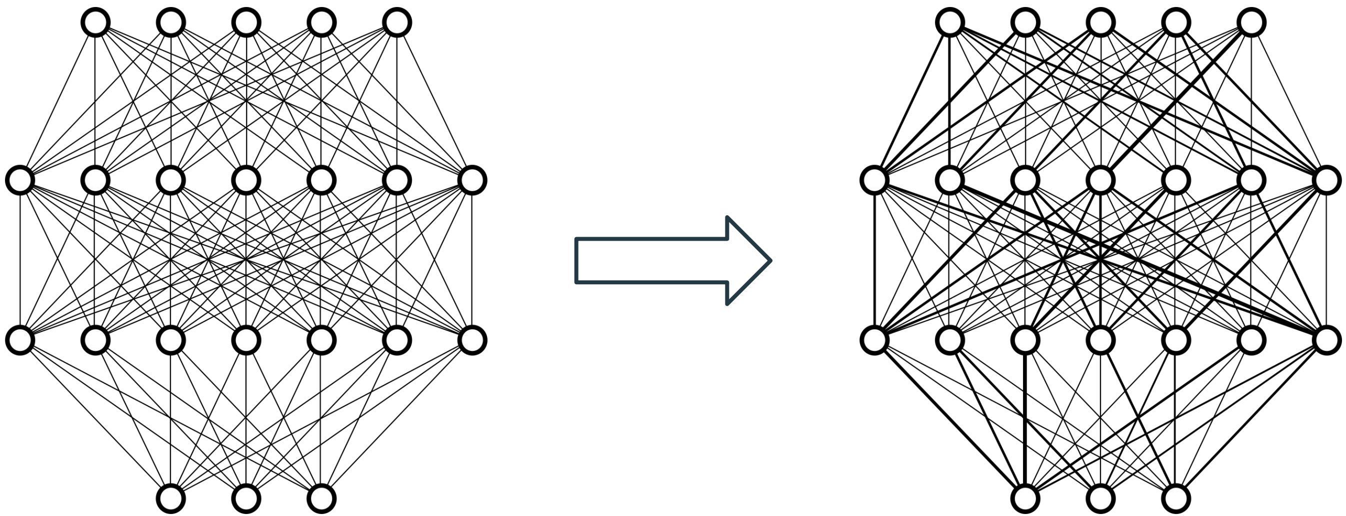
Iterative magnitude pruning (IMP)

[Lottery ticket hypothesis - FrankleCarbin arXiv:1803.03635]



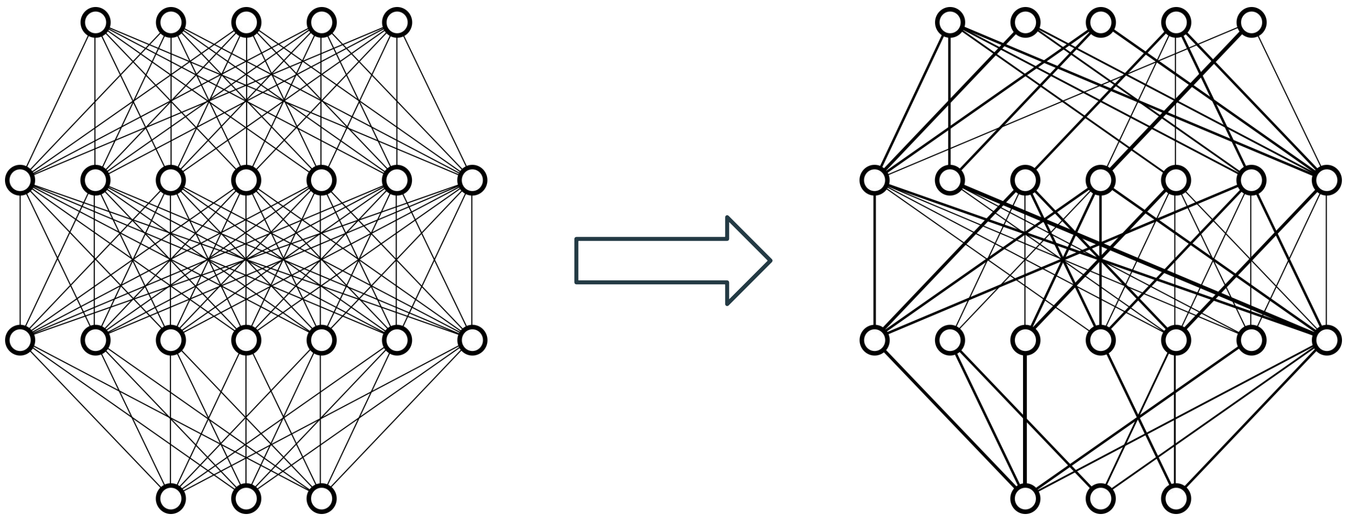
Iterative magnitude pruning (IMP)

[Lottery ticket hypothesis - FrankleCarbin arXiv:1803.03635]



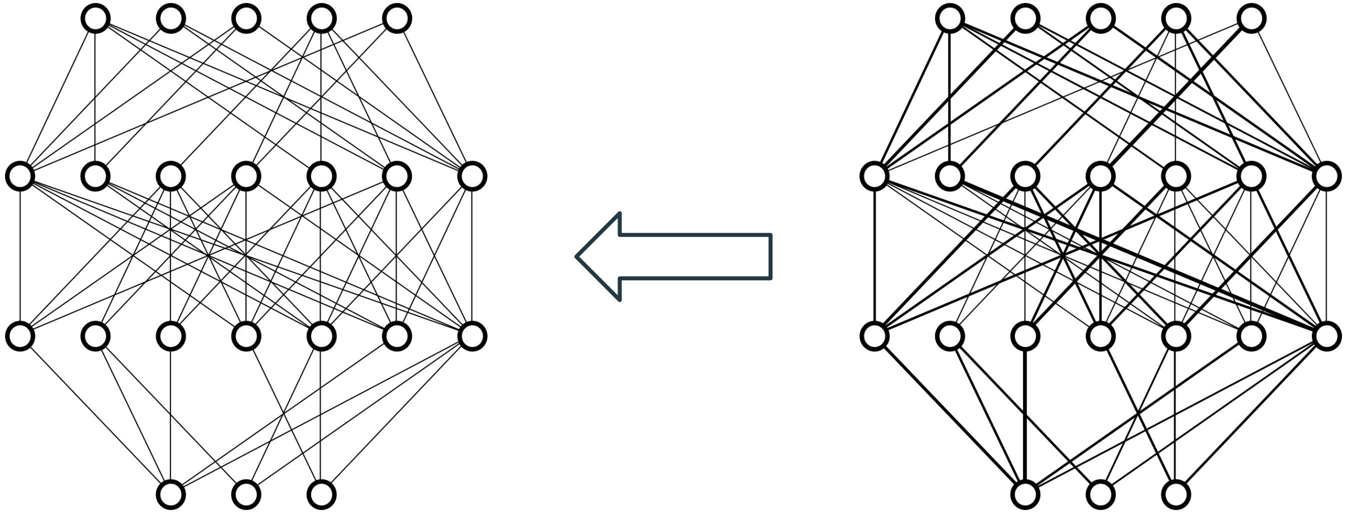
Iterative magnitude pruning (IMP)

[Lottery ticket hypothesis - FrankleCarbin arXiv:1803.03635]



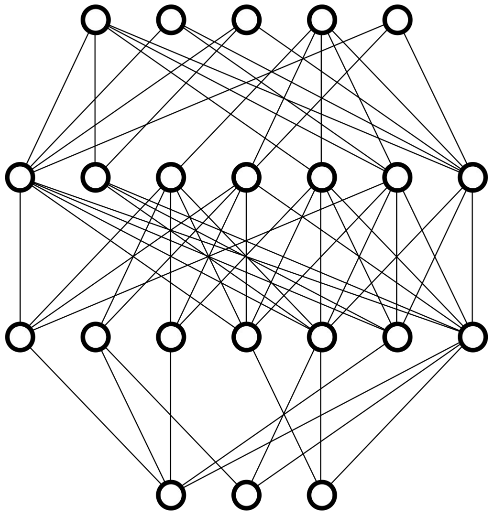
Iterative magnitude pruning (IMP)

[Lottery ticket hypothesis - FrankleCarbin arXiv:1803.03635]



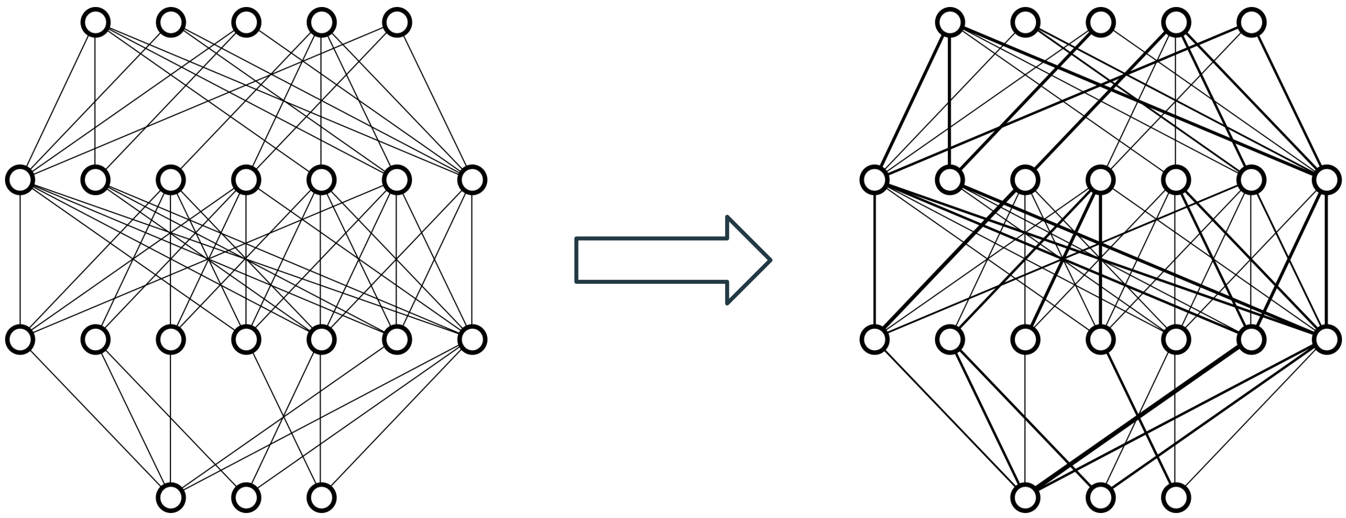
Iterative magnitude pruning (IMP)

[Lottery ticket hypothesis - FrankleCarbin arXiv:1803.03635]



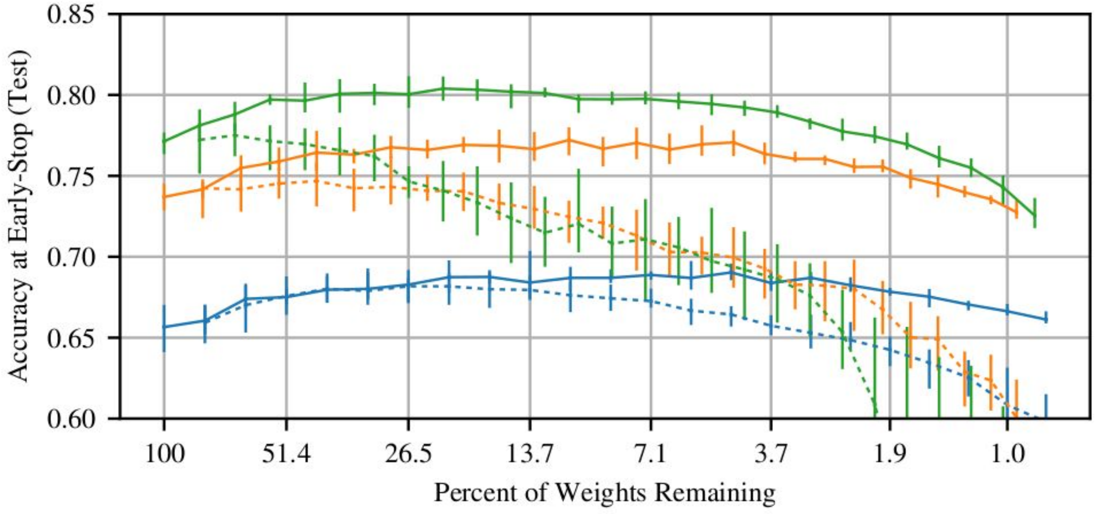
Iterative magnitude pruning (IMP)

[Lottery ticket hypothesis - FrankleCarbin arXiv:1803.03635]



Lottery ticket hypothesis (LTH) [FrankleCarbin18]

The smaller architecture can solve the problem… often with better generalization!

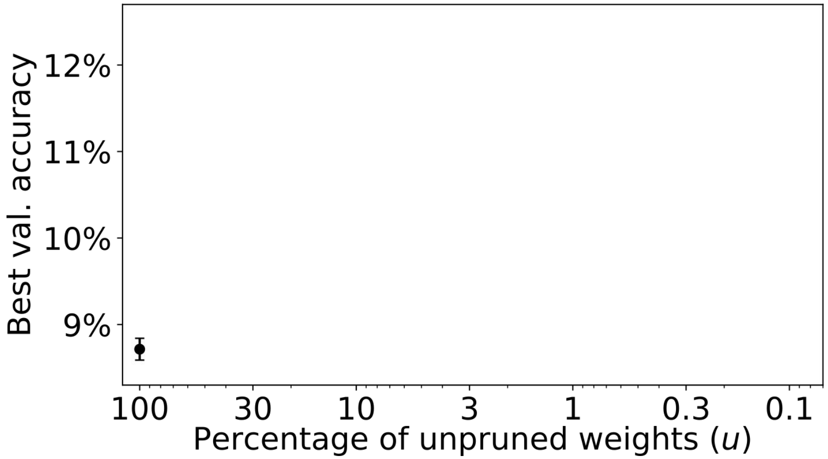


[FrankleCarbin, arXiv:1803.03635, IMP of small CNNs on CIFAR10]

What characterizes these minima that are missed by the FCN training?

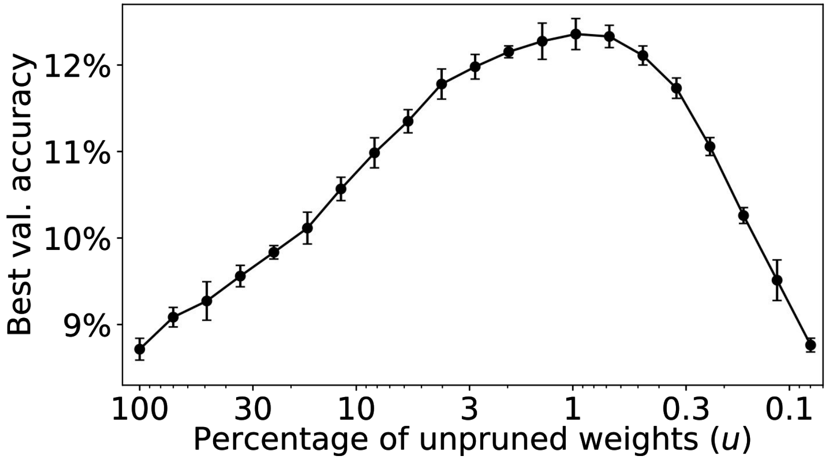
Pruning a simple FCN on ImageNet32

- ImageNet data (1.2M images, 1k classes) scaled to 32x32
- 3 hidden layers [3072:(1024x3):1000], ReLU, cross-entropy
- Iterative Magnitude Pruning (IMP), 30% per iteration



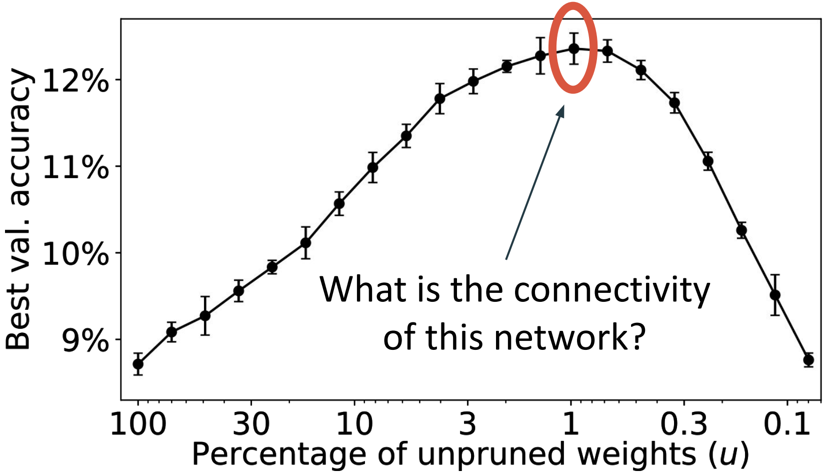
Pruning a simple FCN on ImageNet32

- ImageNet data (1.2M images, 1k classes) scaled to 32x32
- 3 hidden layers [3072:(1024x3):1000], ReLU, cross-entropy
- Iterative Magnitude Pruning (IMP), 30% per iteration



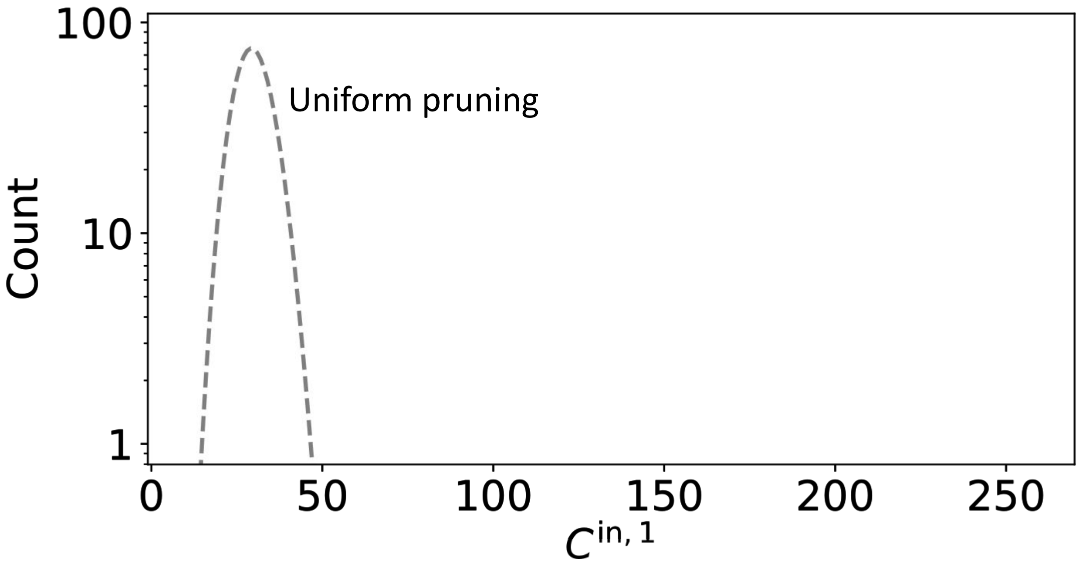
Pruning a simple FCN on ImageNet32

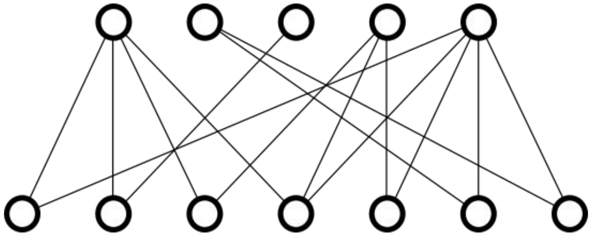
- ImageNet data (1.2M images, 1k classes) scaled to 32x32
- 3 hidden layers [3072:(1024x3):1000], ReLU, cross-entropy
- Iterative Magnitude Pruning (IMP), 30% per iteration



First layer: connectivity

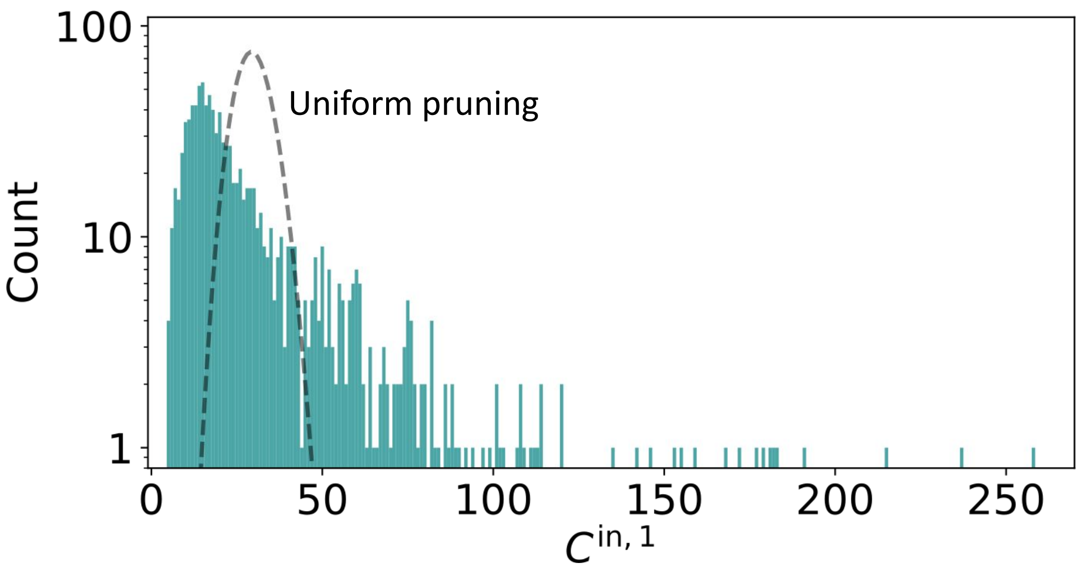
Some nodes retain a large number of connections to the input.

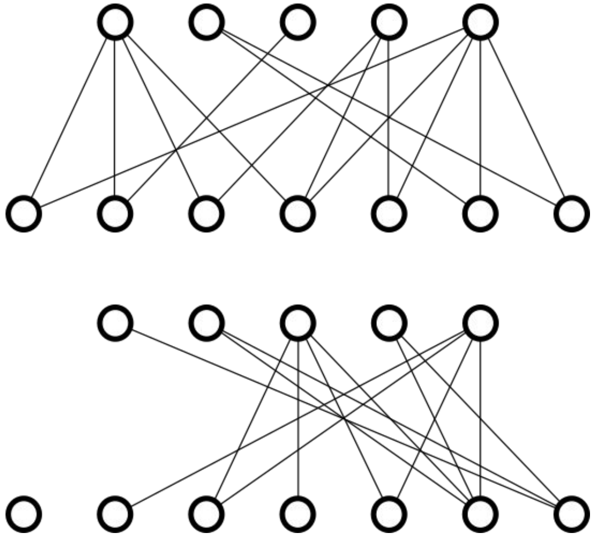




First layer: connectivity

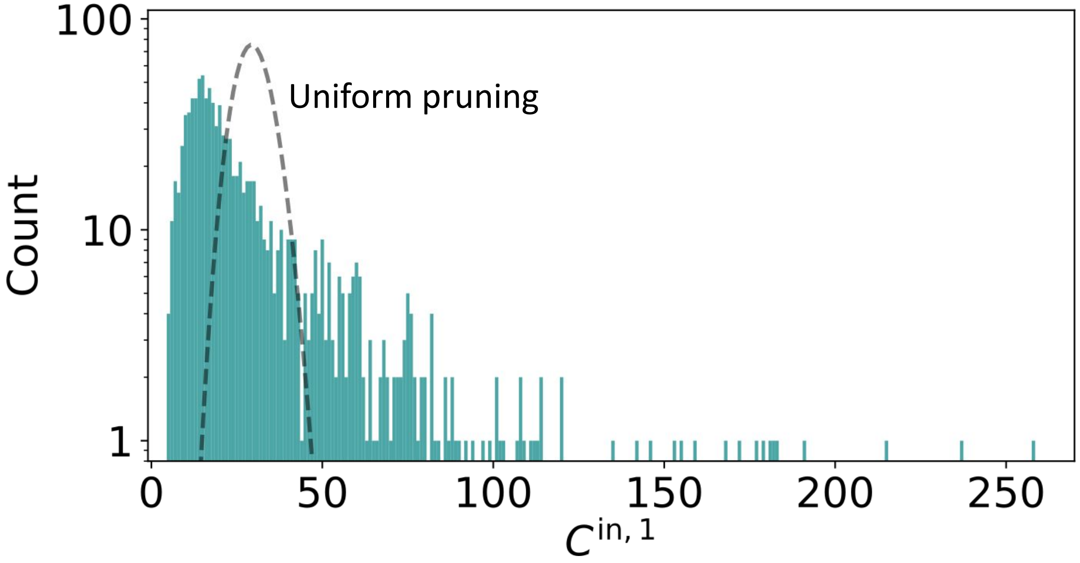
Some nodes retain a large number of connections to the input.

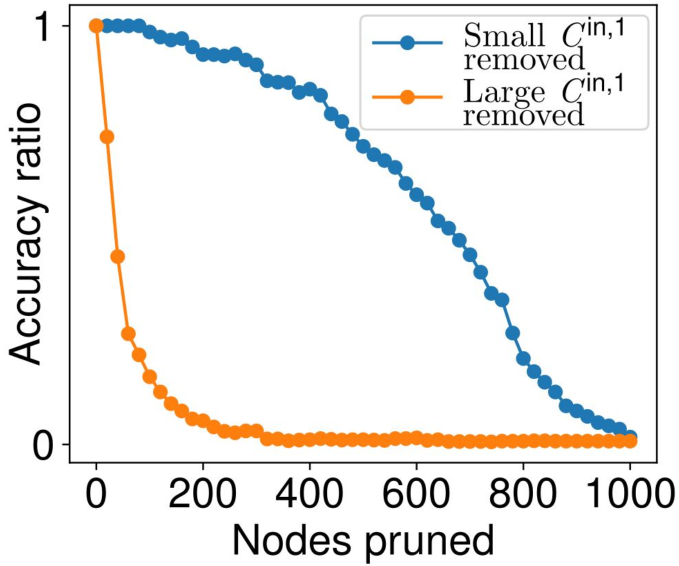




First layer: connectivity

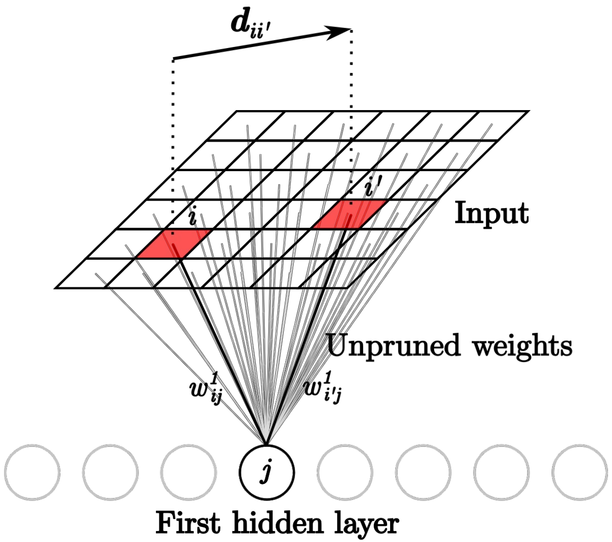
Some nodes retain a large number of connections to the input.





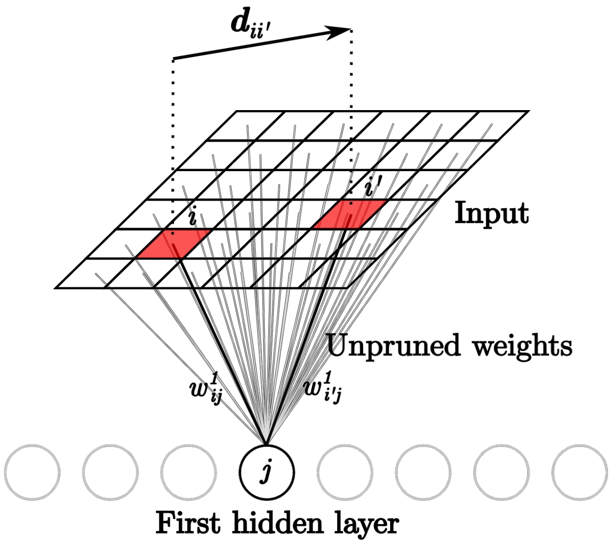
First layer: locality

Relative position of pixels connected to the same hidden node.

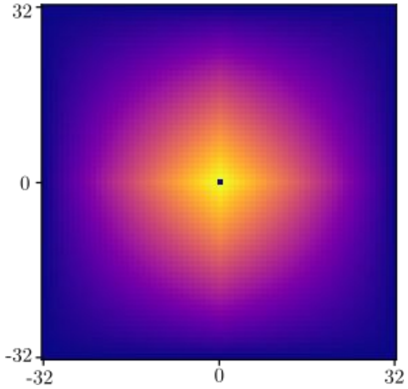


First layer: locality

Relative position of pixels connected to the same hidden node.

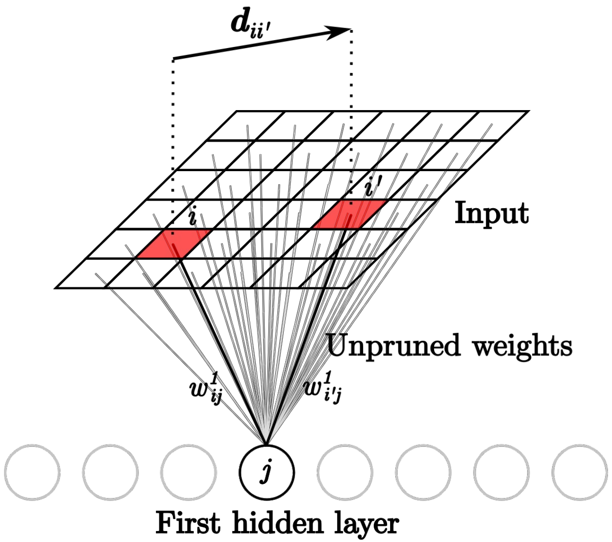


Dense Network

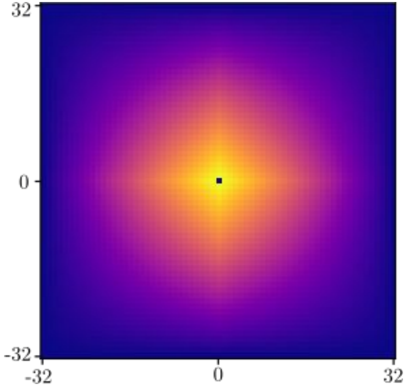


First layer: locality

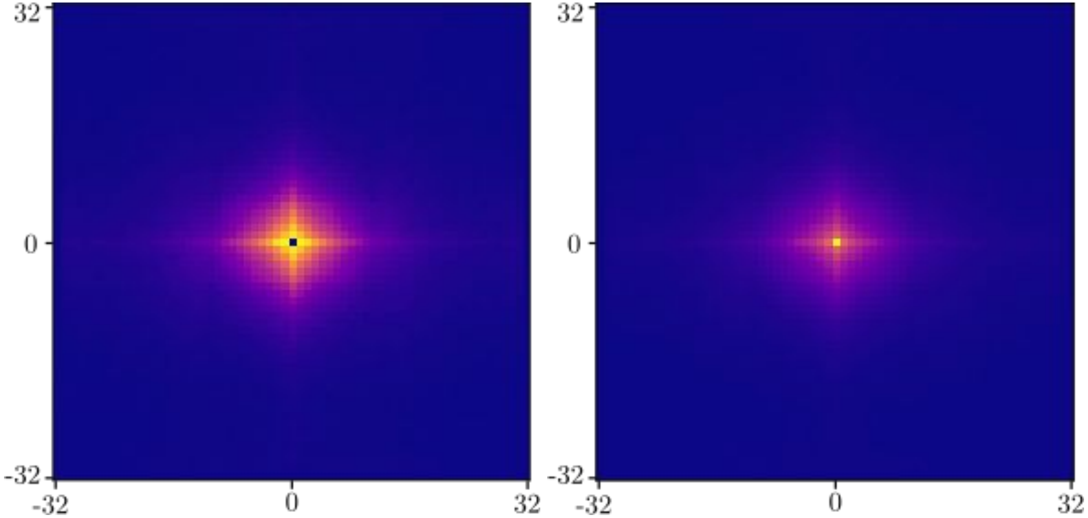
Relative position of pixels connected to the same hidden node.



Dense Network



Pruned Network

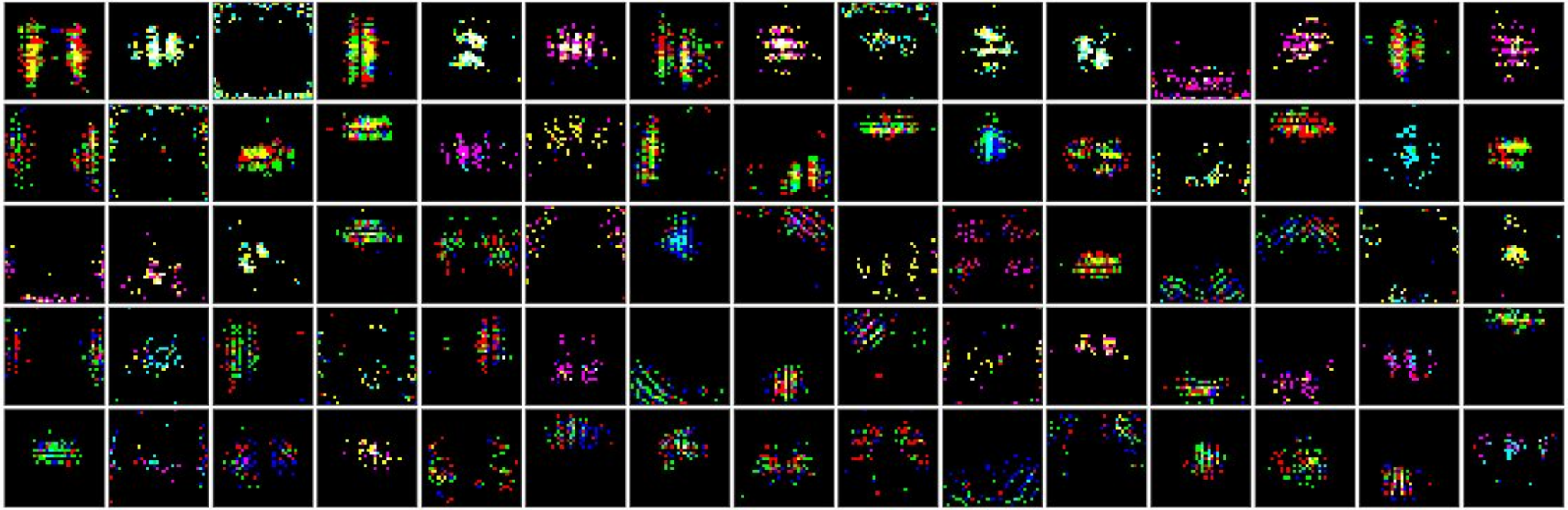


Same color channel

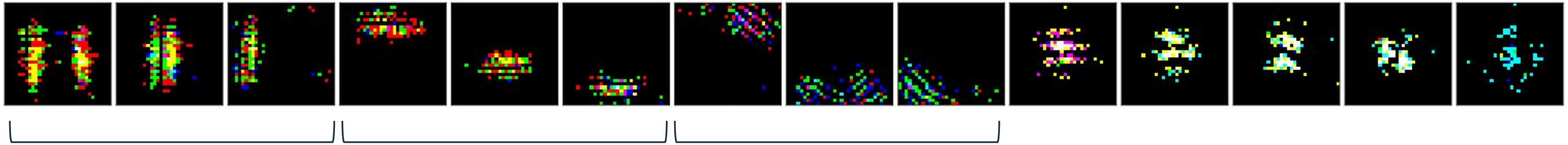
Different color channels

First layer: masks are structured

Most connected nodes show local and highly structured patterns.

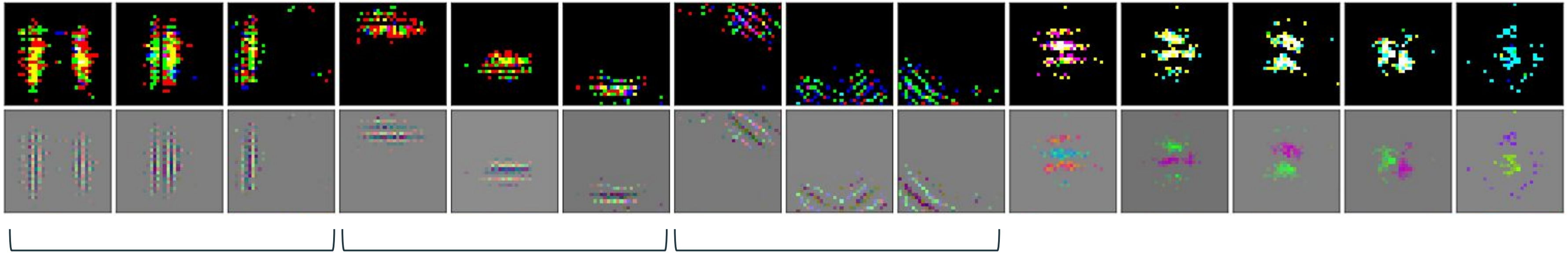


First layer masks: translational invariance



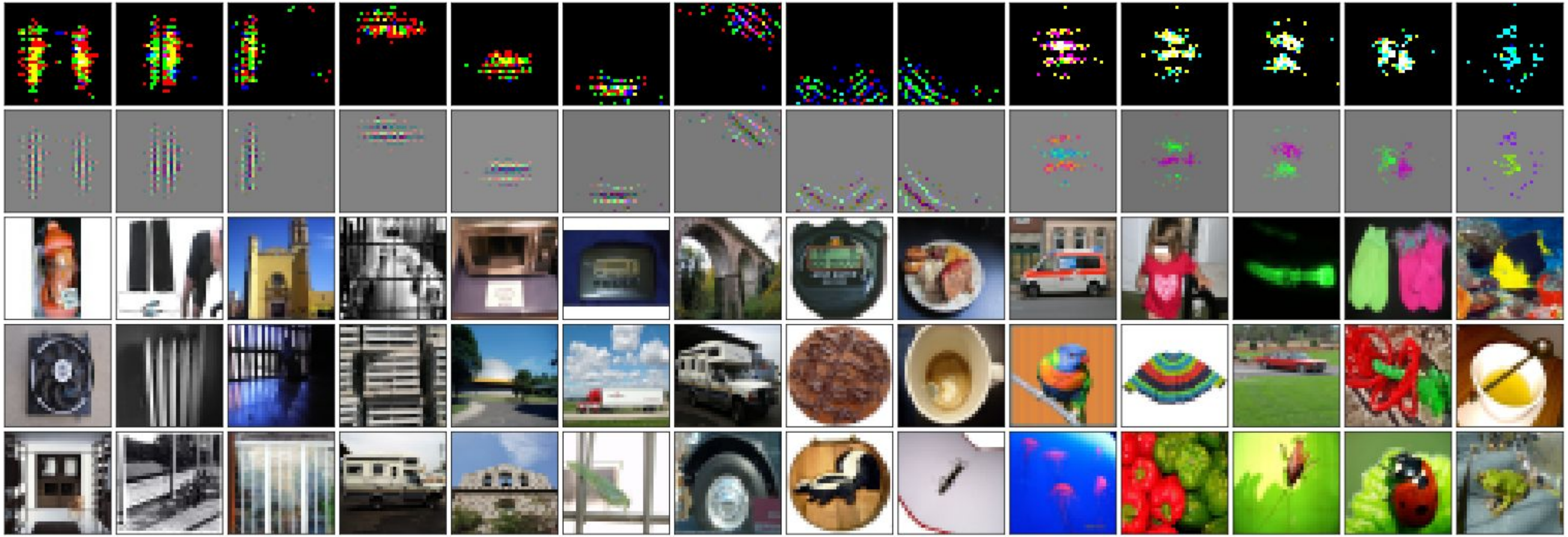
Similar patterns at different locations

First layer masks: translational invariance



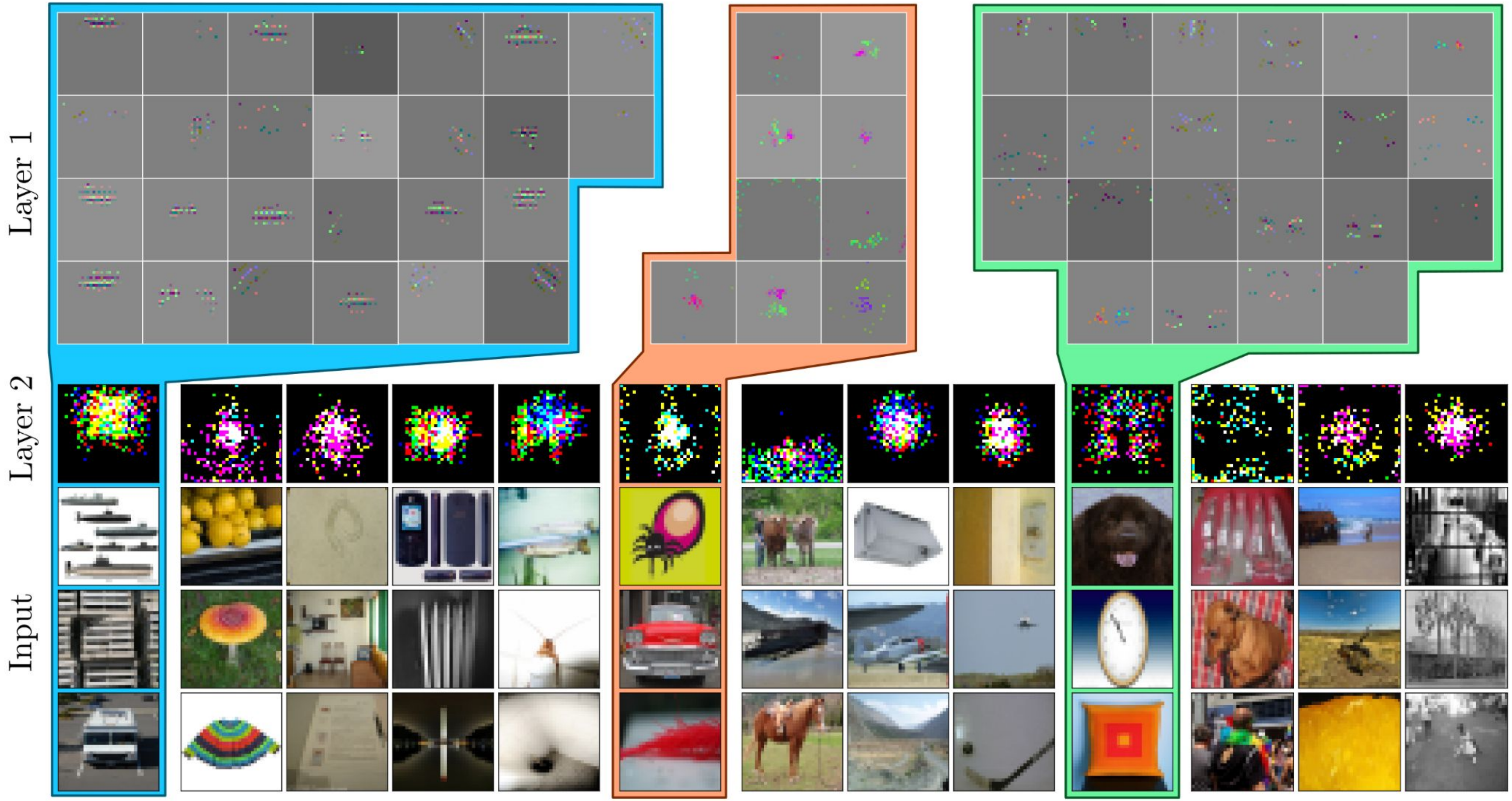
Similar weights

First layer masks: translational invariance

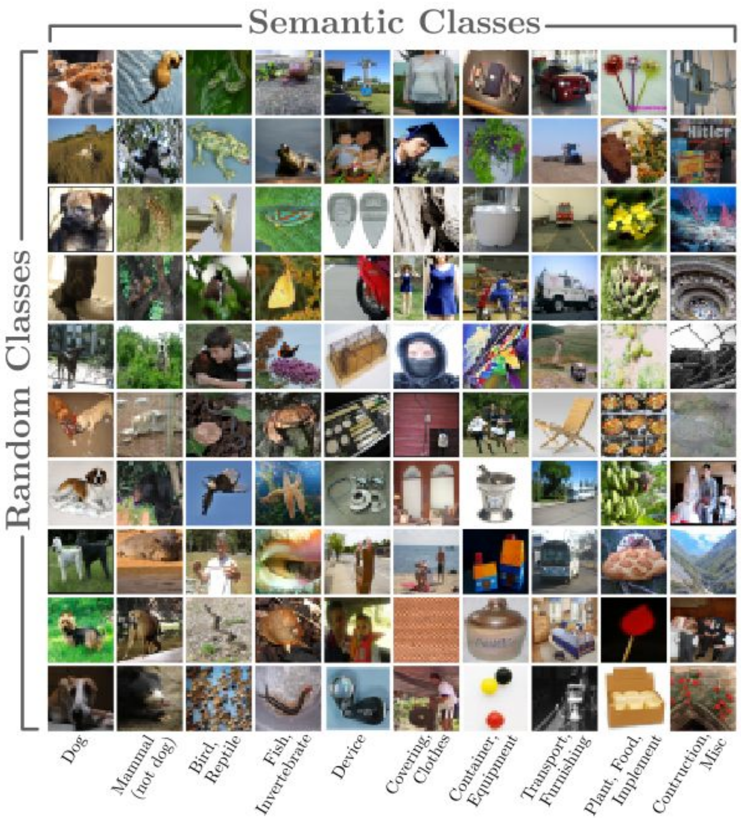


Class independent response

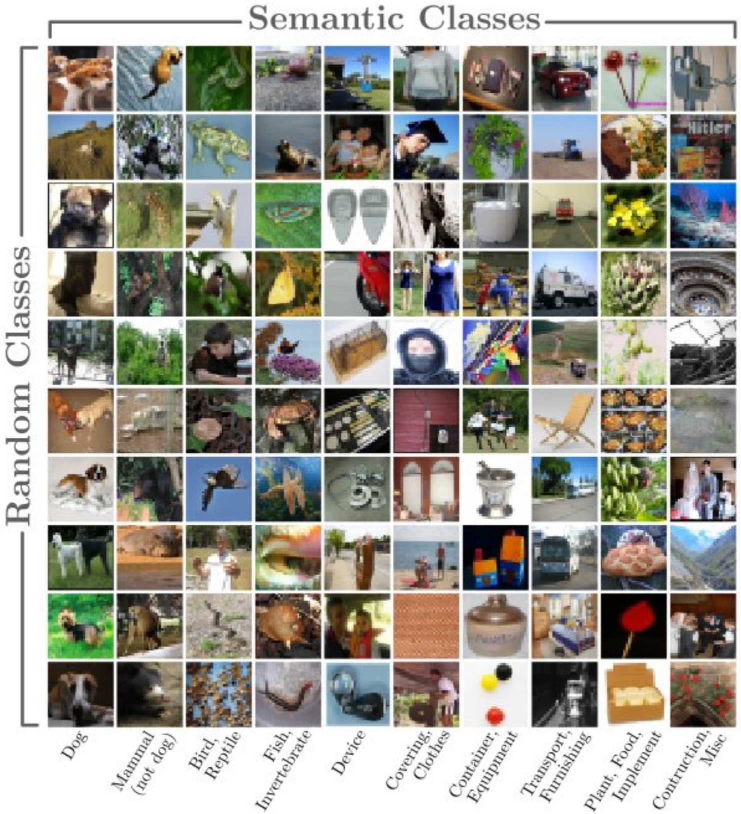
Second layer: aggregated features

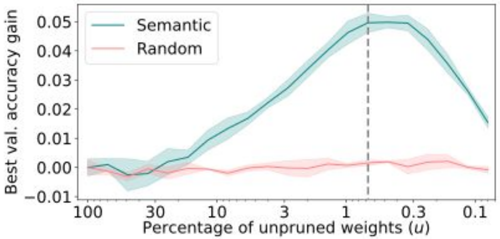


Varying the task: 10 superclasses

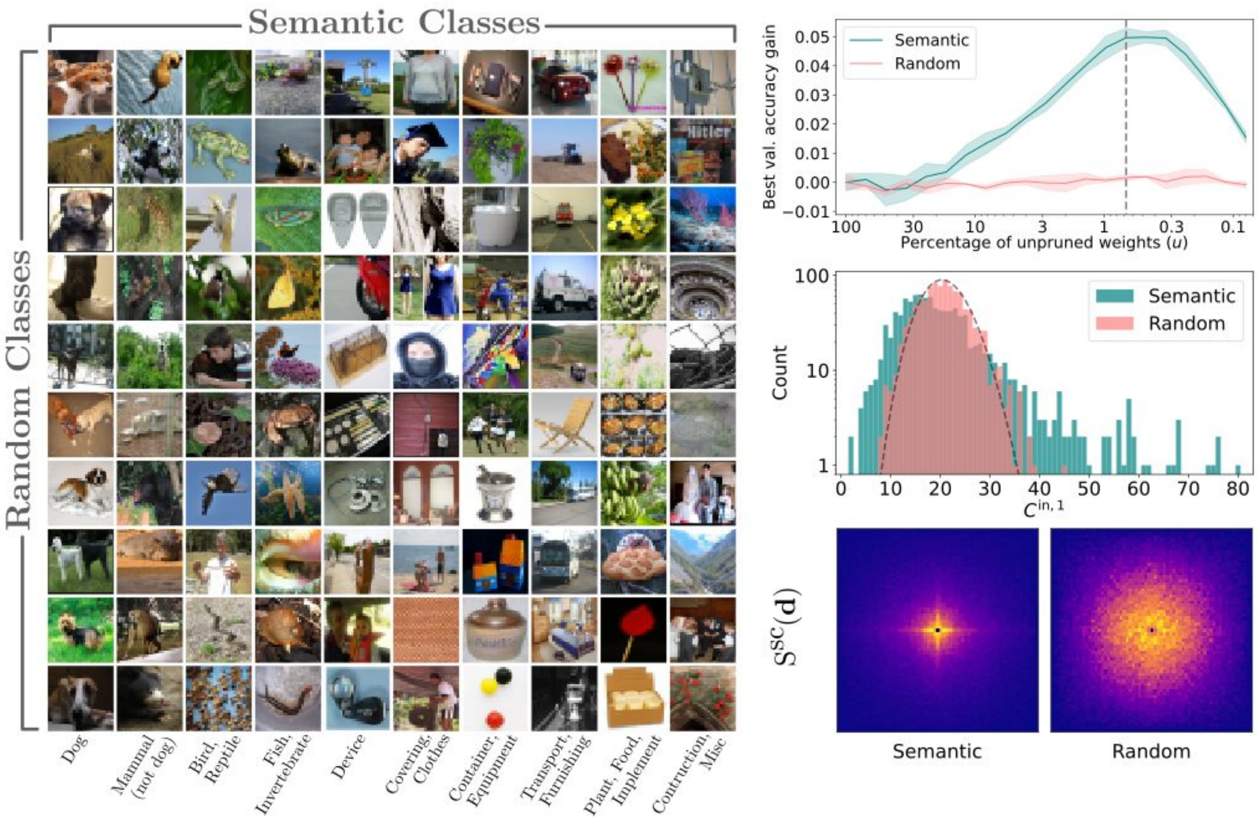


Varying the task: 10 superclasses

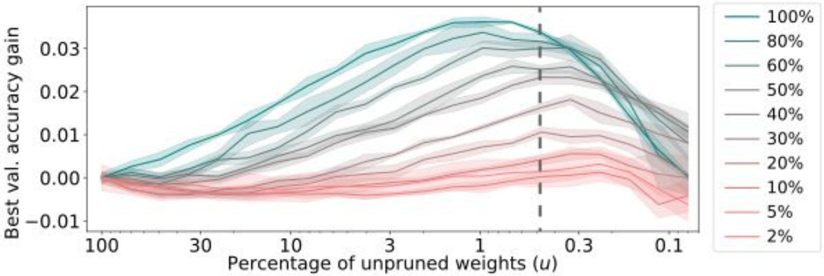


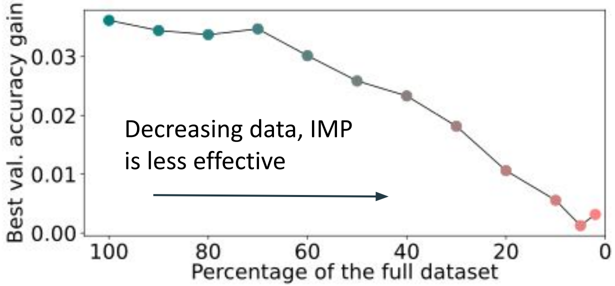


Varying the task: 10 superclasses

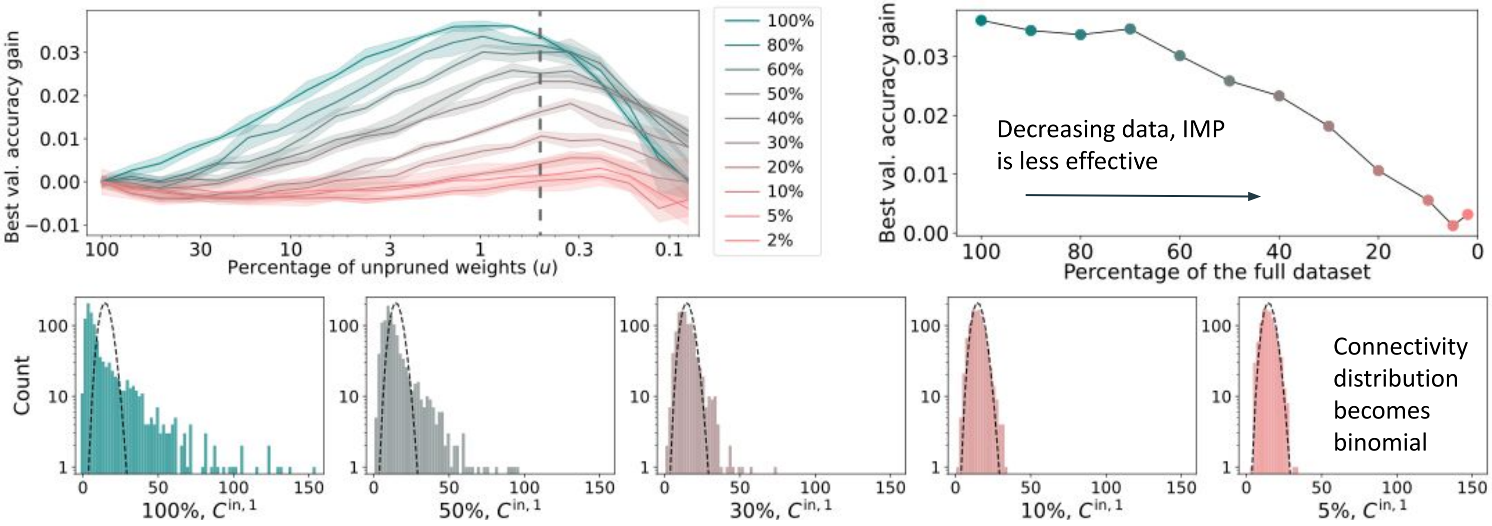


Decreasing the size of the dataset

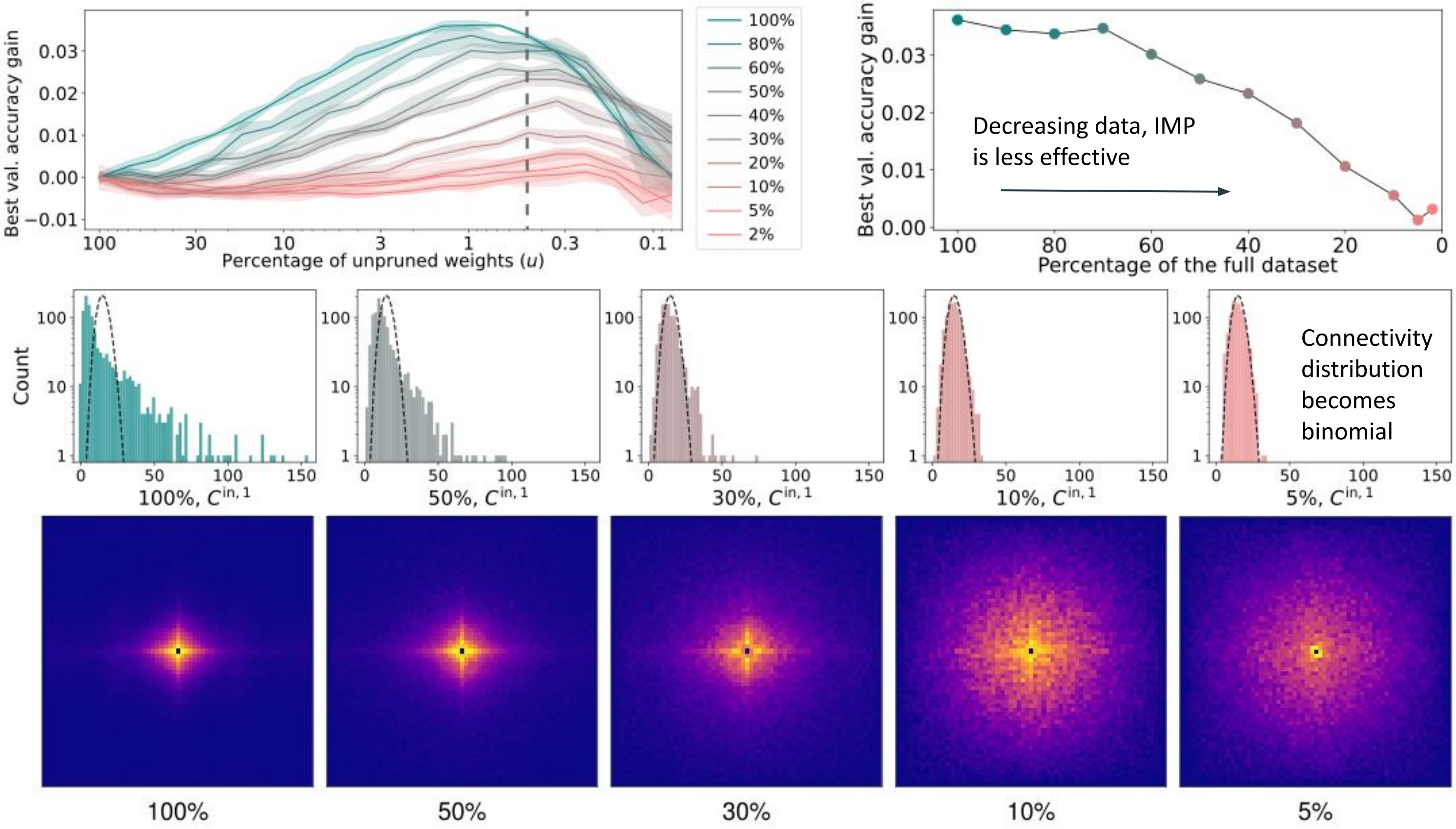




Decreasing the size of the dataset



Decreasing the size of the dataset



Locality is lost

Conclusions

- IMP on images leads to masks that are
  - local
  - structured
  - hierarchical
- The structure is related to success of the task
- IMP can help uncover architectural bias

Thank you!