Triangular Dropout:

Variable Network Width without Retraining

ICML Workshop on Dynamic Neural Networks
July 22, 2022

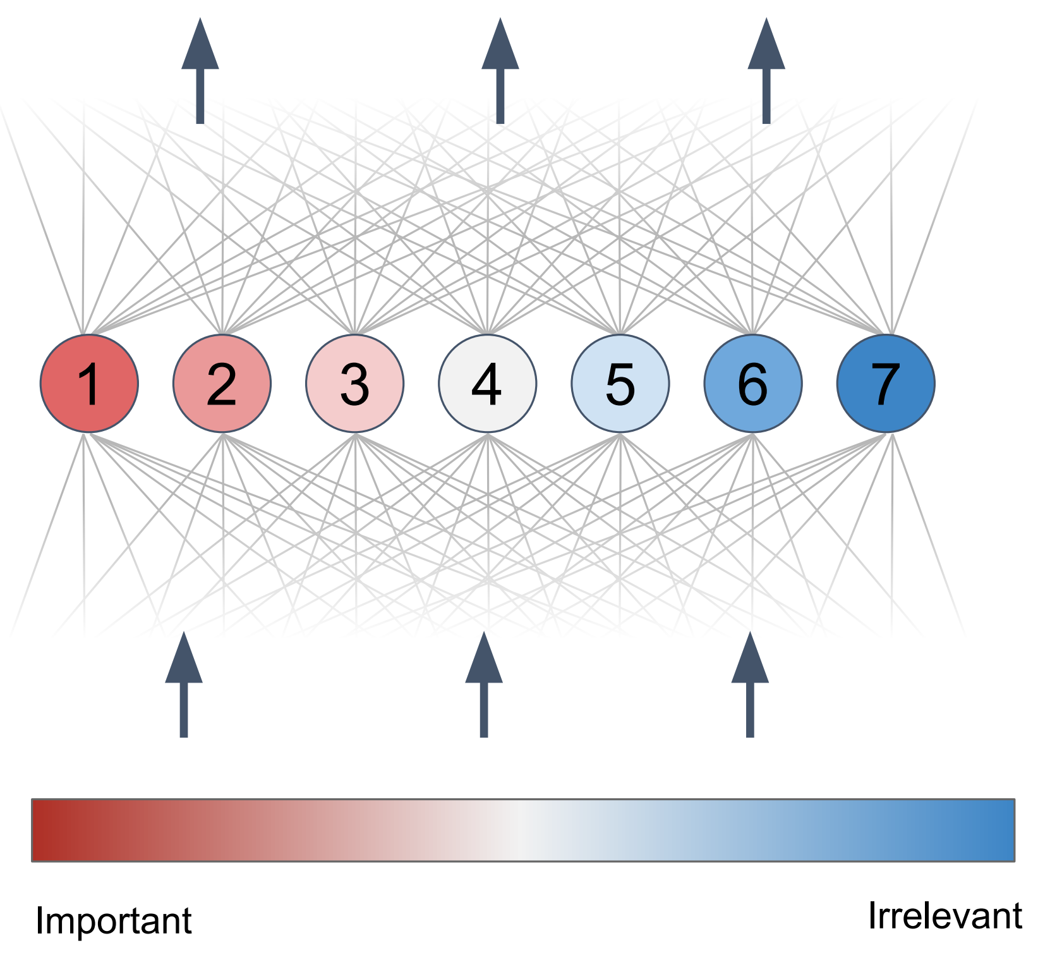
Edward W. Staley
Jared Markowitz
Johns Hopkins University
Applied Physics Lab (APL)

What is Triangular Dropout?

Triangular Dropout is a neural network training technique that leads to fully-connected layers with two useful properties:

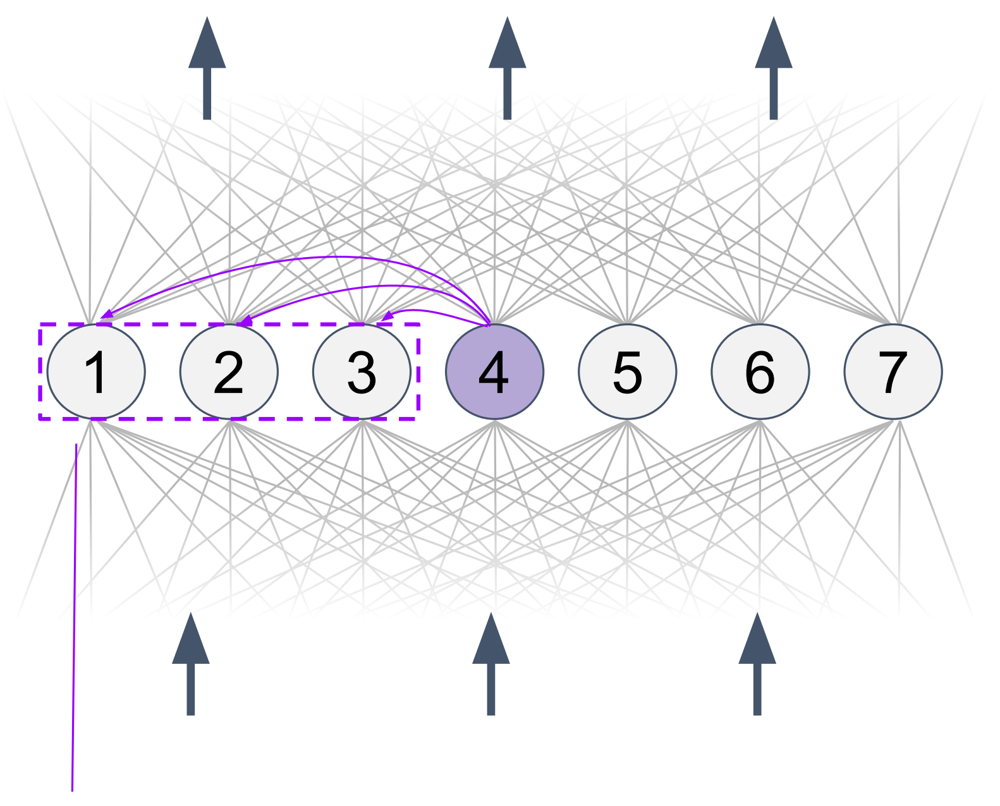
(1) Organized Feature Importance

Learned features (layer outputs) are organized by importance. The most important high-level features form on the left, while less critical features form on the right.



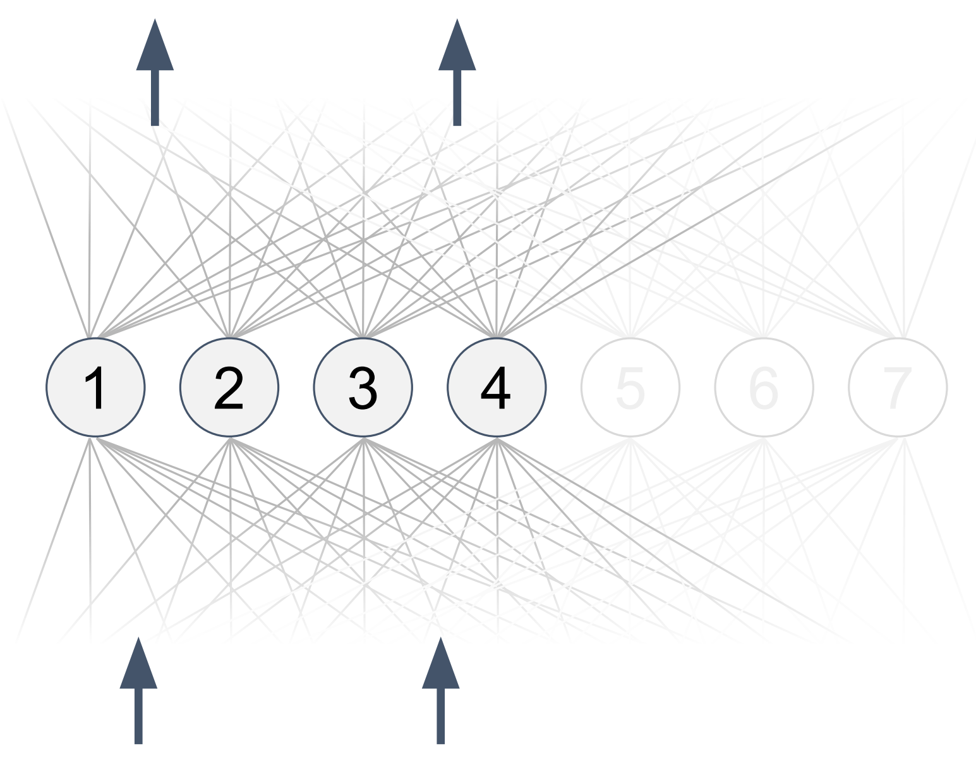
(2) Asymmetric Feature Dependence

Learned features are less interdependent than in a typical network. Instead, a given feature mainly depends on those features which are more important than itself.



Feature 4 mainly depends on these features.

These properties result in a fully-connected layer which can be densely pruned <u>after</u> training, resulting in a **variable-width layer**.

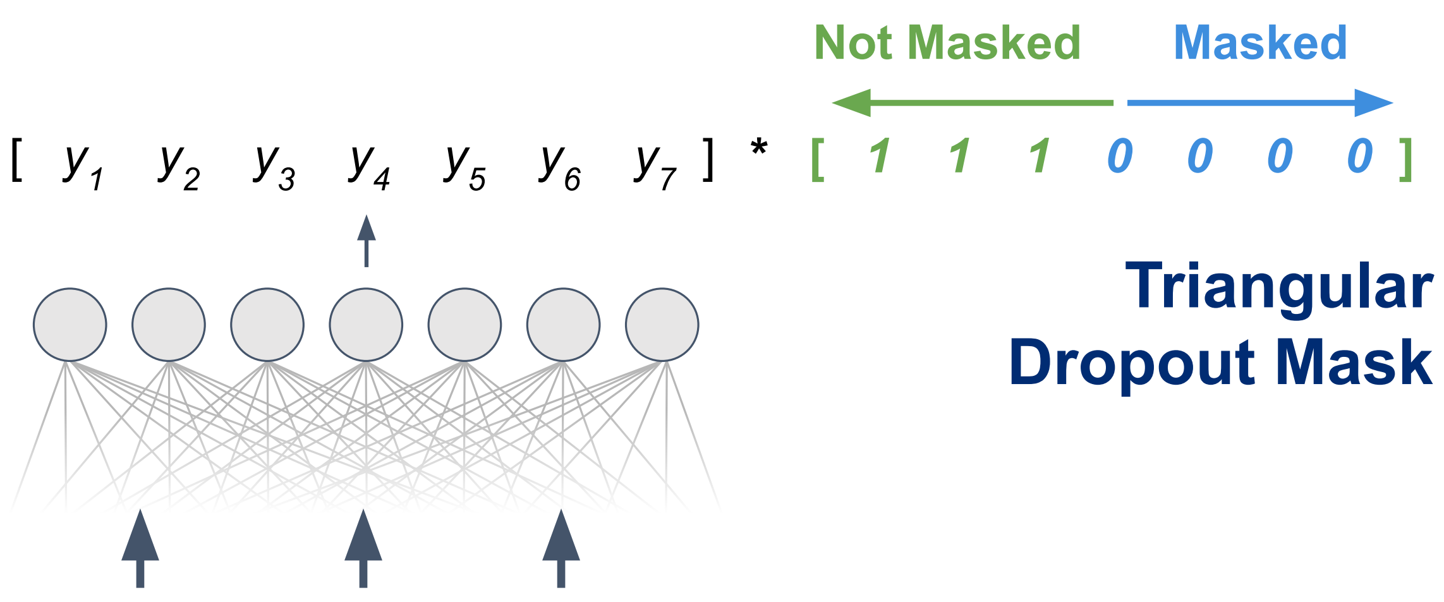


- More important features on the left can function without the existence of less important features on the right.
- Pruning away less important features trades performance for number of parameters.

How does it work?

Triangular Dropout is very simple to apply in practice.

Modifying standard dropout, we use binary masks which are not fully random. They consist of 1's followed by 0's:



$[\ y_1\ \ y_2\ \ y_3\ \ y_4\ \ y_5\ \ y_6\ \ y_7\ ] * [\ 1\ \ 1\ \ 1\ \ 0\ \ 0\ \ 0\ \ 0\ ]$

Triangular Dropout Mask

This results in masked layer outputs in which a given feature always co-exists with other features to the left, but not the right. This creates the two properties discussed in the first panel.

$$[\ y_1\ \ y_2\ \ y_3\ \ 0\ \ 0\ \ 0\ \ 0\ ]$$

Feature 3 can be learned jointly with [1,2], but not [4-7]

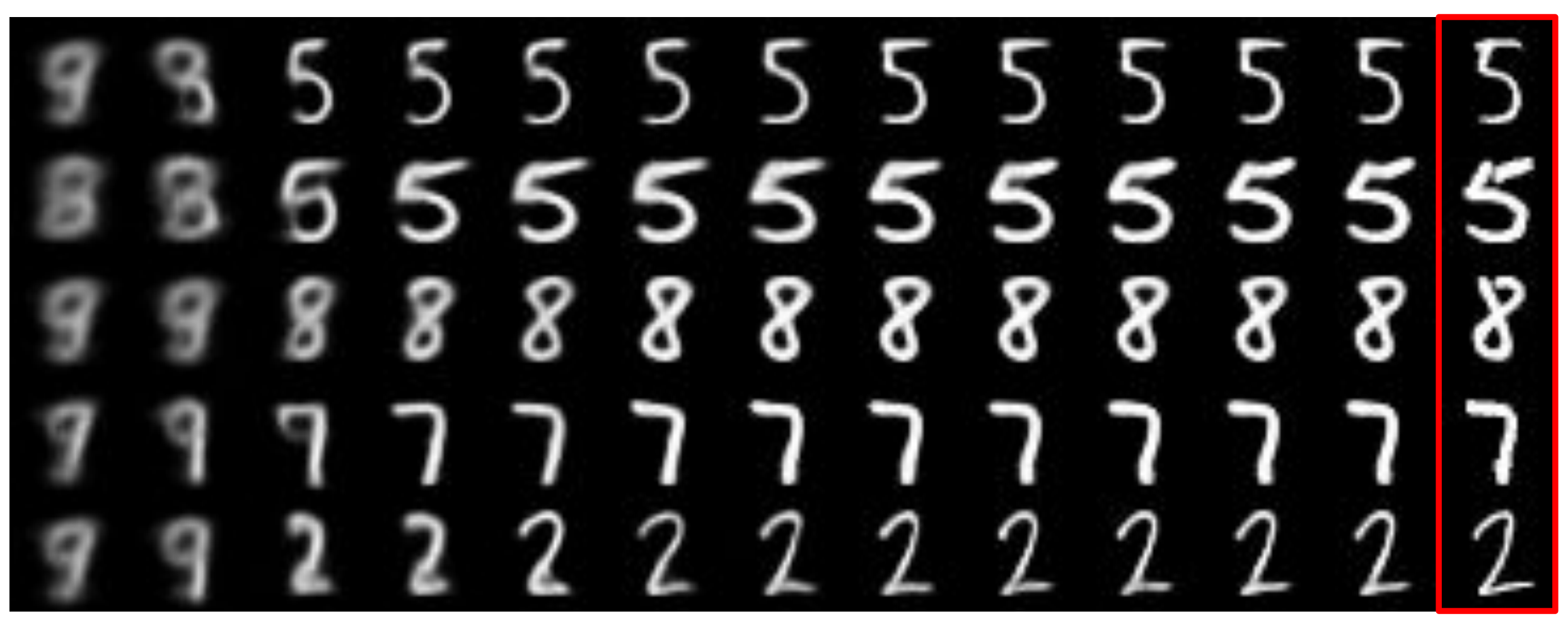
In practice, we enumerate all such masks over the batch, which is simply a lower triangular matrix that can be applied to the entire batch at once. Thus, we named our technique **Triangular Dropout.**

$$\begin{array}{c}
[\ 1\ \ 0\ \ 0\ \ 0\ \ 0\ \ 0\ \ 0\ ]\\
[\ 1\ \ 1\ \ 0\ \ 0\ \ 0\ \ 0\ \ 0\ ]\\
[\ 1\ \ 1\ \ 1\ \ 0\ \ 0\ \ 0\ \ 0\ ]\\
[\ 1\ \ 1\ \ 1\ \ 1\ \ 0\ \ 0\ \ 0\ ]\\
[\ 1\ \ 1\ \ 1\ \ 1\ \ 1\ \ 0\ \ 0\ ]\\
[\ 1\ \ 1\ \ 1\ \ 1\ \ 1\ \ 1\ \ 0\ ]\\
[\ 1\ \ 1\ \ 1\ \ 1\ \ 1\ \ 1\ \ 1\ ]
\end{array}$$

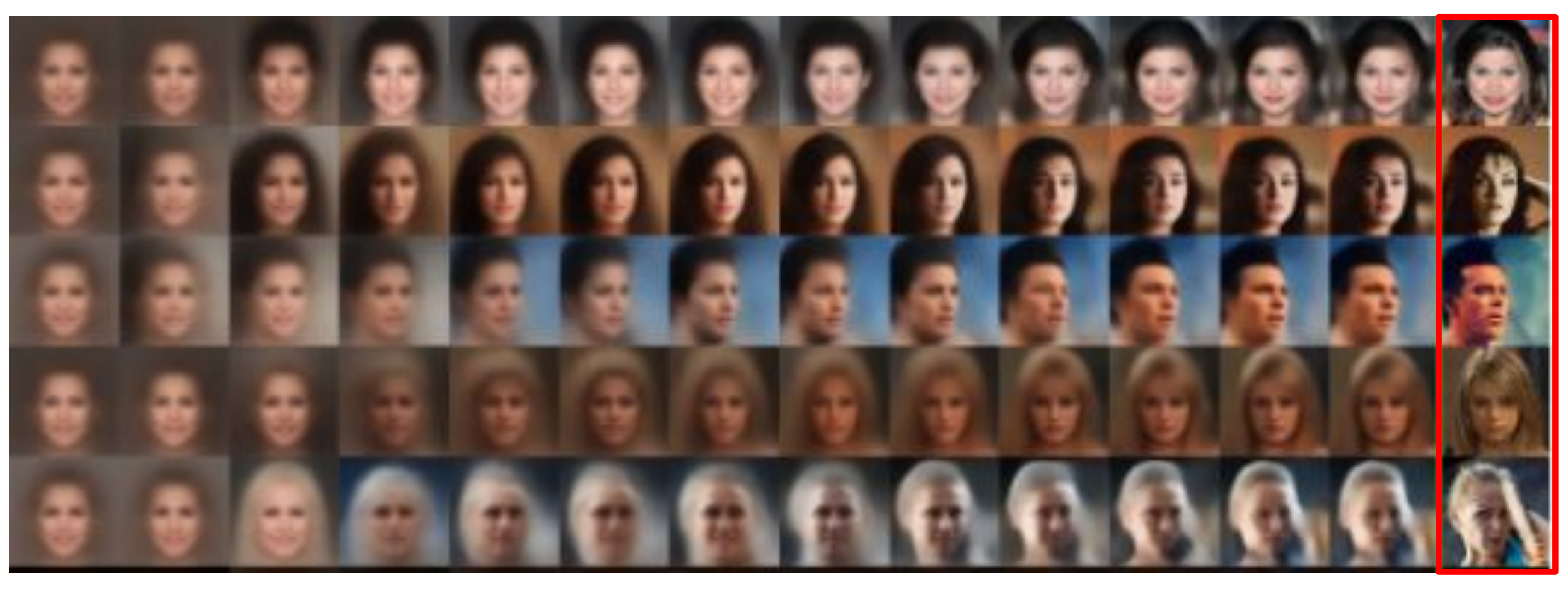
Application: Variable-Sized Autoencoders

If Triangular Dropout is applied to the encoding layer of an autoencoder, it results in an organized encoding that can be arbitrarily pruned to select a more compressed encoding.

After only training the autoencoder once, we can remove less important features from the encoding and see the fidelity of reconstructions decrease. See two examples below:



MNIST Ground Truth



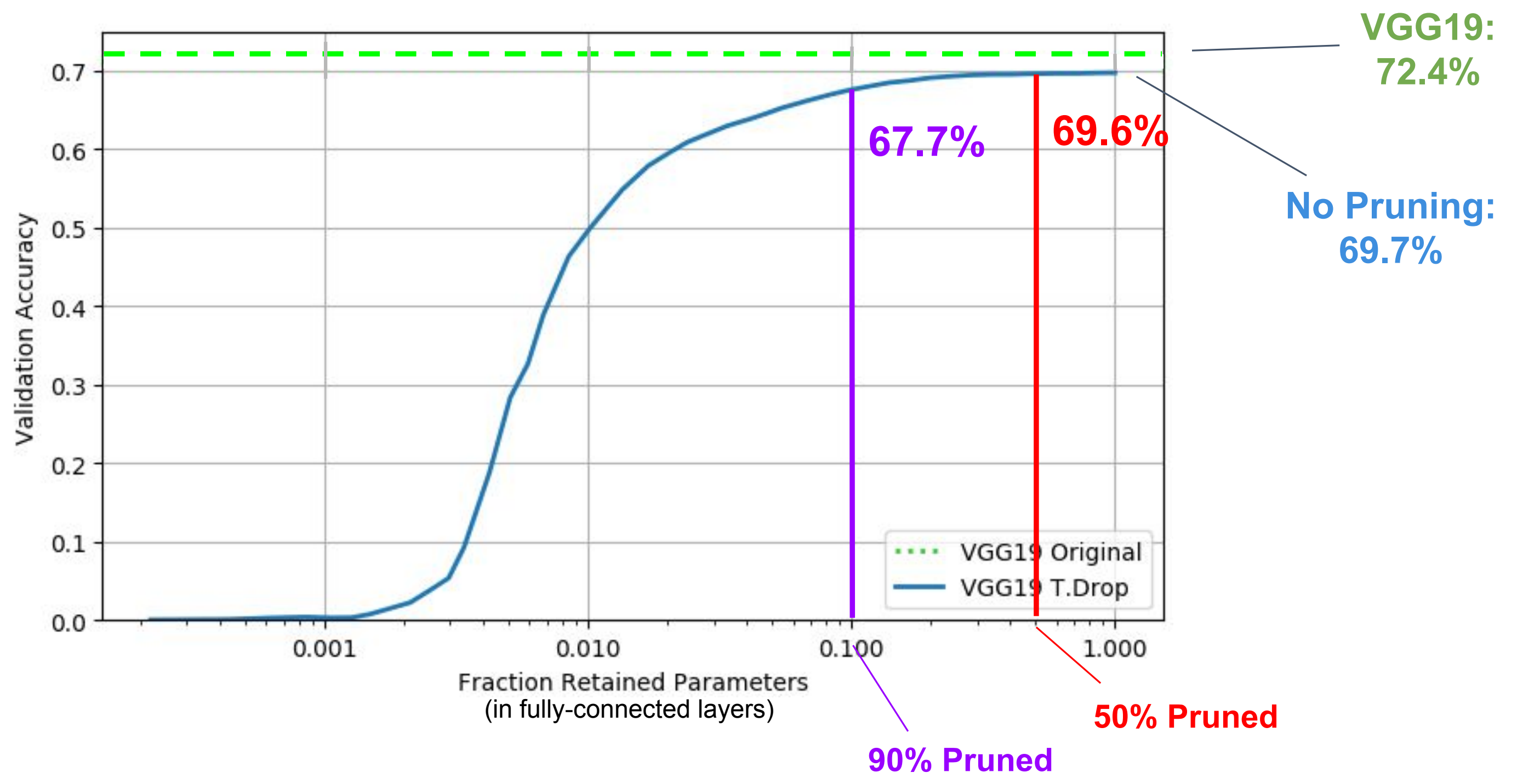
CelebA Ground Truth

More Compressed ⟷ More Features

Application: Overparameterization

In an overparameterized architecture, Triangular Dropout can be used to extract a functional, dense sub-network after training.

Here we consider the fully-connected portion of VGG19 on ImageNet, which is known to be overparameterized. After retraining these layers with Triangular Dropout, we can vastly reduce their width with only minimal impact to our accuracy.



Note that our network does not reach the original performance of VGG19, presumably because Triangular Dropout imparts some constraints on the network during training. Additionally, we have only developed our method for MLPs, and thus could not apply it to the convolutions. We leave these items to future work.

Correspondence to:
Edward W. Staley (edward.staley@jhuapl.edu)
Jared Markowitz (jared.markowitz@jhuapl.edu)

This work relates to Department of Navy award N00014-20-1-2239 issued by the Office of Naval Research. The United States Government has a royalty-free license throughout the world in all copyrightable material contained herein. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the Office of Naval Research.