Split-Brain Autoencoders

Lekha Iyengar, Diego Casabuena, Aishwarya Budhkar

Local Context
Domain Gap

Autoencoders
Hinton et al. 2006

+Denoising
Vincent et al. 2008

+Stacked
Vincent et al. 2010

+Convolutional
Masci et al. 2011

Global Context
Input Handicap

Denoising and Impainting
Xie et al. 2012

Context Encoding
Pathak et al. 2016

Cutout
DeVries et al. 2017

Cross-Channel Predictions
Useful Representation

Colorization
Zhang et al. 2016

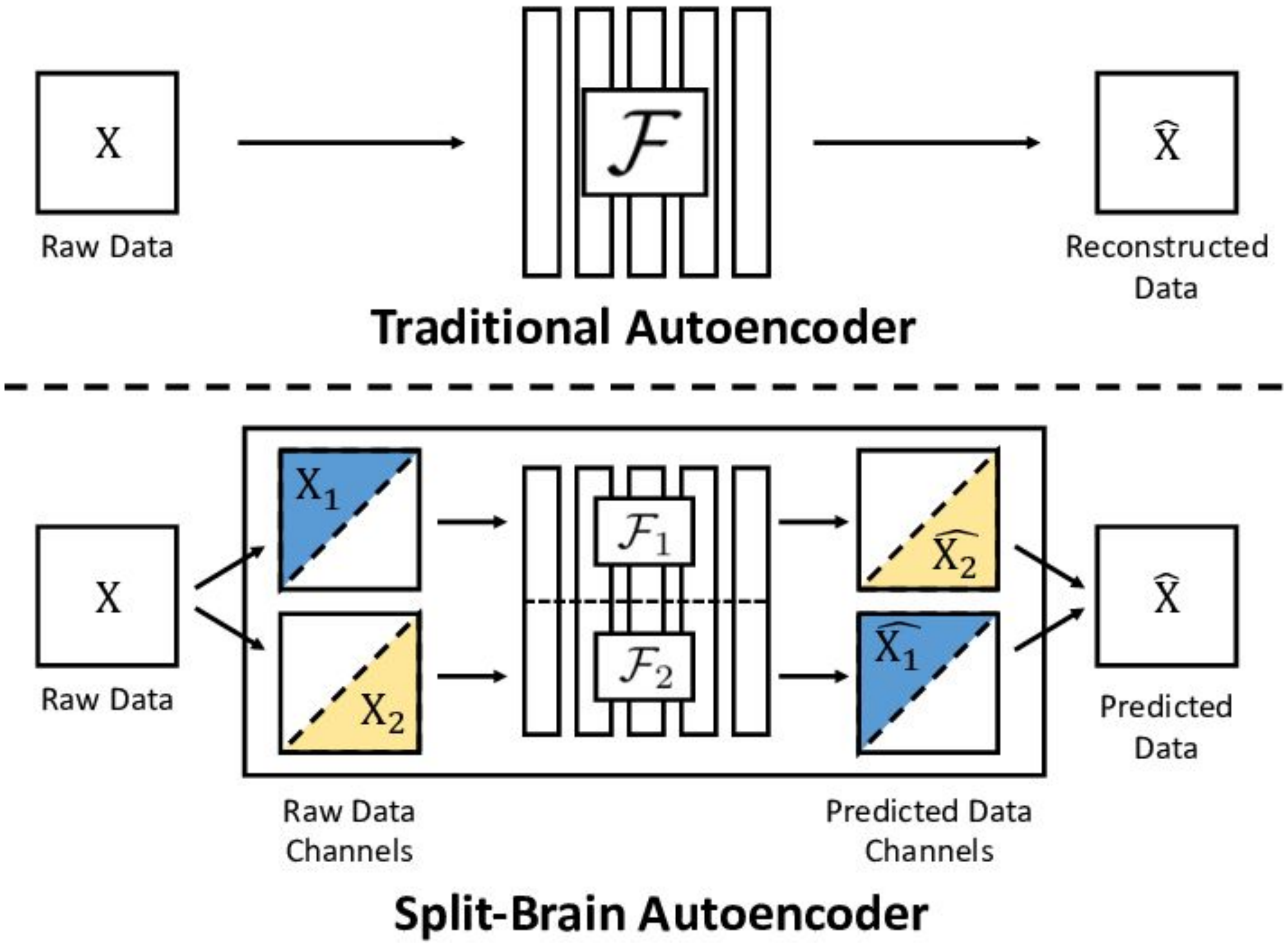
Understanding from Colors
Larsson et al. 2017

Automatic Colorisation
Larsson et al. 207

Split Brain
Zhang et al. 2017

Self-Supervised Pretraining from Colorization

1. Represent image in a color space that separates luminosity and color.
2. Separate image channels accordingly.
3. Pass each subset through disjoint fully convolutional architecture to predict the other subset.
4. Obtain a downsampled (12x12) and quantized version of the input.
5. Take the Cross Entropy Loss pixelwise w.r.p. the downsampled image.
6. Finetune by adding a classifier as the last layer.

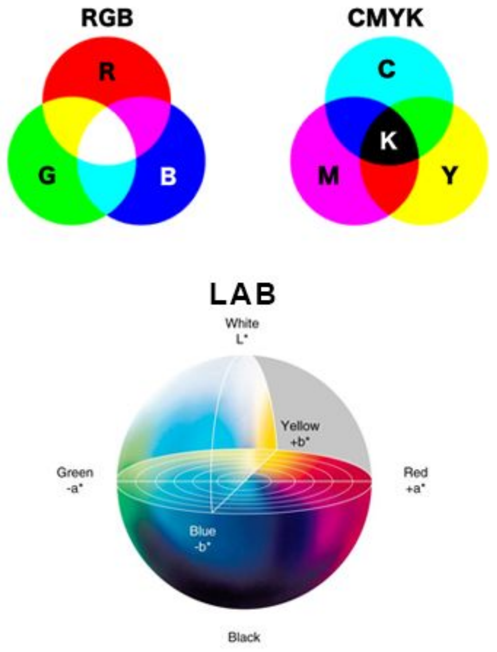


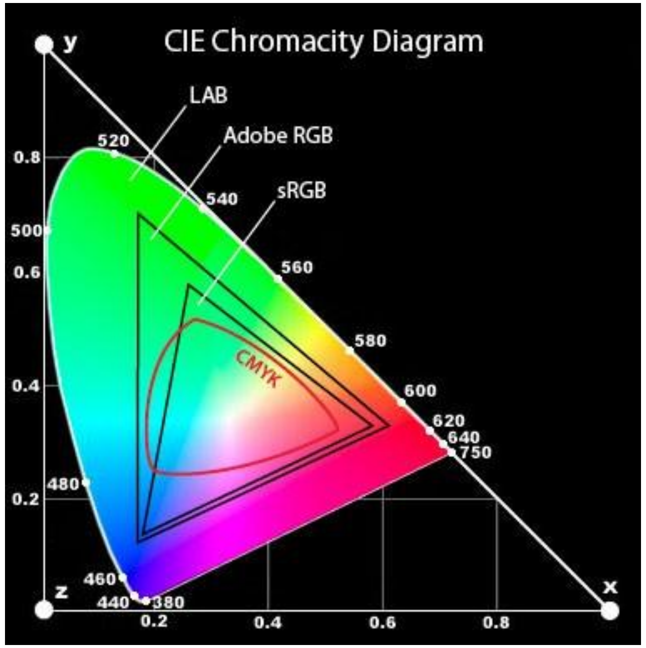
Split-Brain Autoencoders: Unsupervised Learning by Cross-Channel Prediction, Zhang et al. 2017

CIE-Lab Space

- Natural way to represent color.
- Allows for Split Brain to perform colorisation and gray-scale prediction for high level feature extraction.
- RGB->LAB: Naturally ocurring colors are usually within a small subdomain of LAB so the quantization is lossy.
- **Our solution:** Rescale the quantization bins with every image so that reconstruction task learns color contrasts.

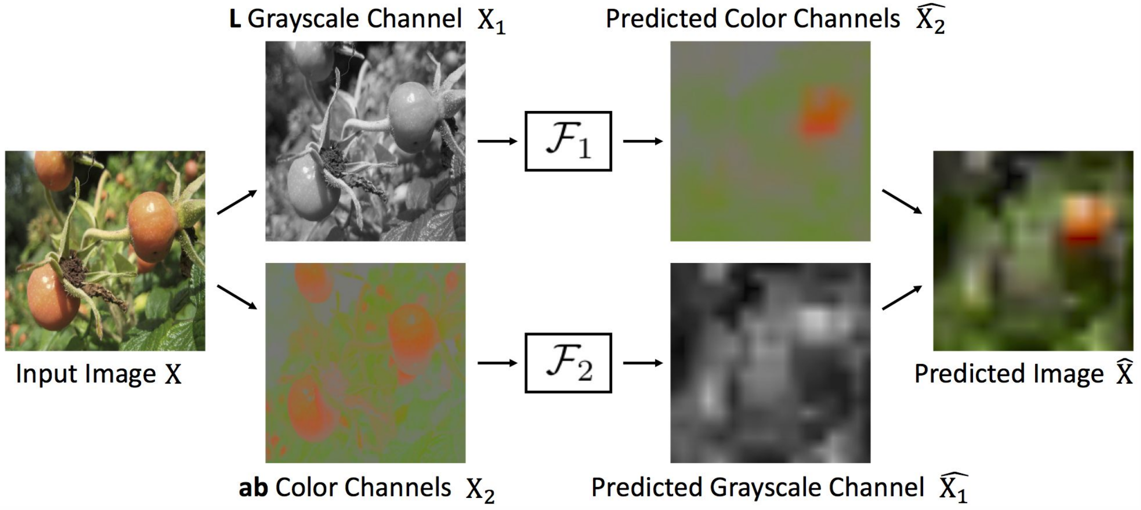
i.e. We care about feature extraction, not image reconstruction.





Color Space Representations (Left), Gamut of Color Spaces (Right), Split Brain Illustration (Bottom)

Original Image (Top), Downsampled Image (Middle), Rescaled Reconstruction (Bottom)



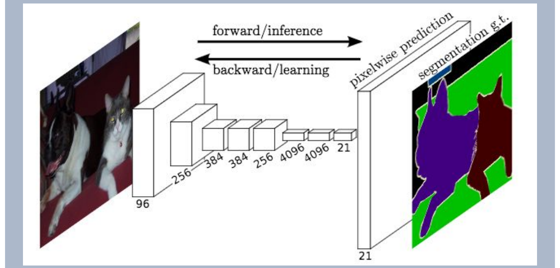
Experiments

Architectures

SimpleNet
(5-layer ConvNet)

FC AlexNet (Below)

FC ResNet 18

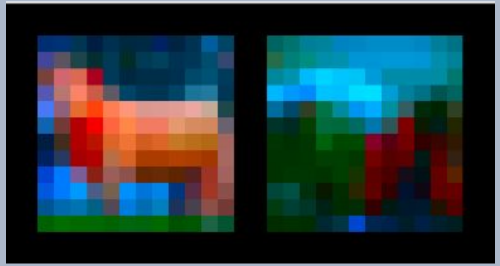


Color Spaces

RGB

CIE-LAB

Re-scaled CIE-LAB
(Our Contribution)



Hyper-parameters

Downsample Size: **12x12**, 16x16, 25x25

Channel 1 Num. Bins
100, 200

Channel 2 Num. Bins
10, 16, 25

Initial Learning Rate
1e-3,**1e-4**,1e-5

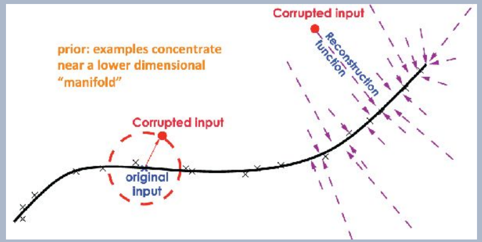
LR Decay
0.1, 0.5, 0.9

Baseline Model

Denoising
Autoencoder

Noise Type:

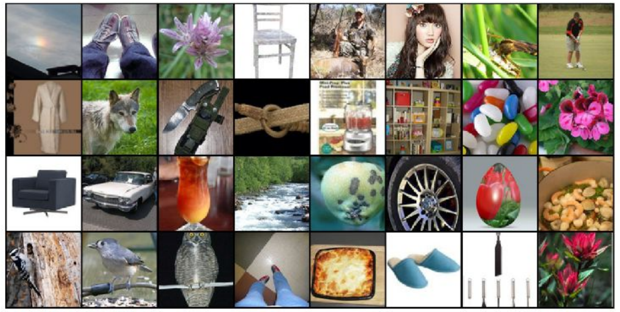
Masking
Salt-and-Pepper
Gaussian



Results

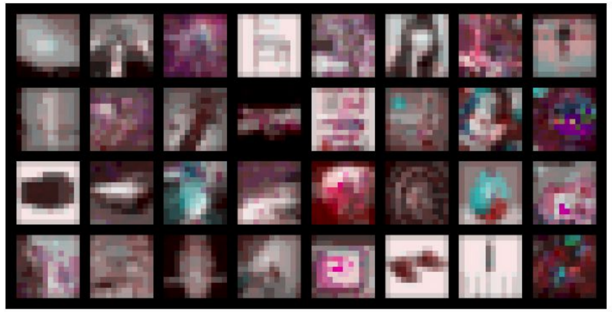
Pre-training Reconstructions ResNet18 (11 Epochs)

Original



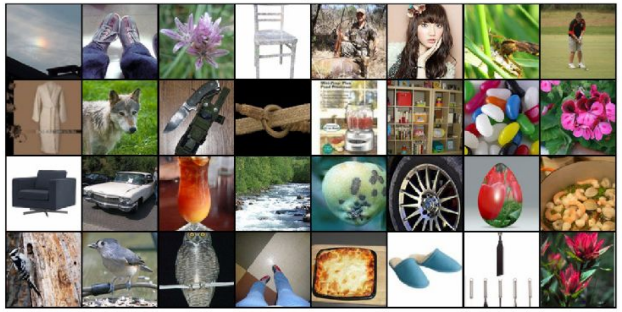
Original Downsized

Reconstructed

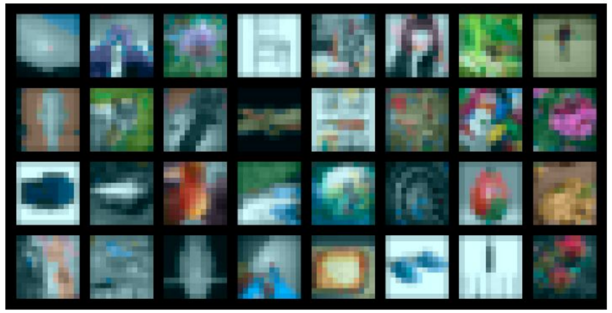


RGB

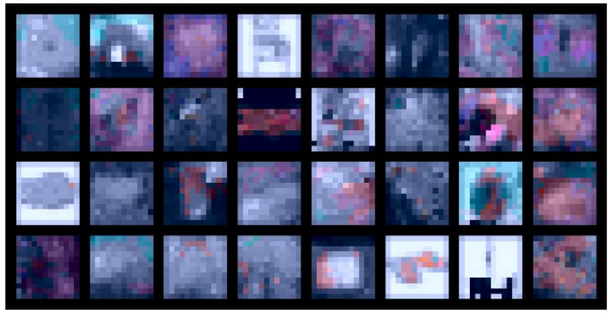
Original



Original Downsized



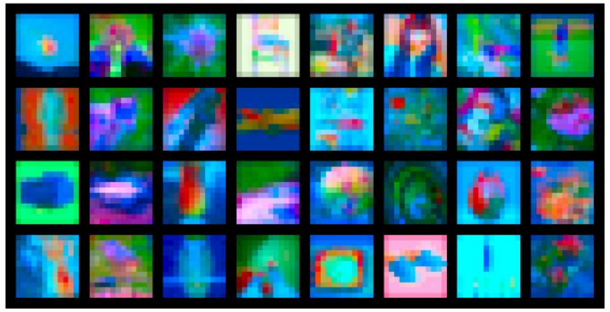
Reconstructed



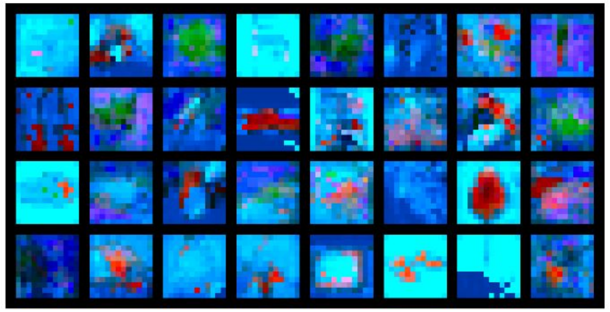
LAB (Split-Brain)

Original

Original Downsized

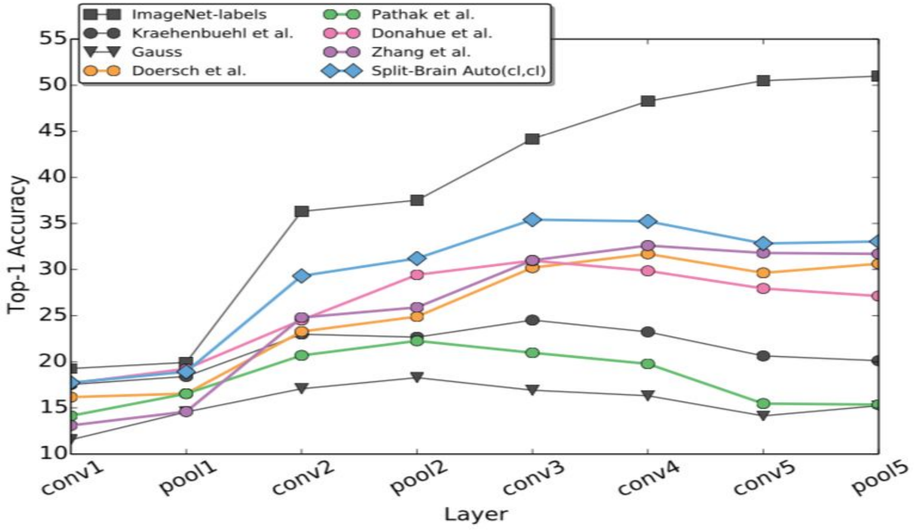


Reconstructed



Re-scaled LAB (Ours)

Original Paper



(a) **ImageNet Classification**

- 1.3M Unlabelled Images (ImageNet)
- Image Size 180x180 (Re-sized ImageNet)
- 40 Epochs of Pre-training
- Best Image Space: CIE-LAB
- Best Model: FC AlexNet

Up to 35% Top-1 Accuracy on ImageNet Classification

Us

Architecture	Image Space	FT Epochs	Top 1 Acc.	Top 5 Acc.
FCAlexNet	RGB	2	0.0943	0.2343
FCAlexNet	LAB	2	0.0301	0.0988
FCAlexNet	Re-scaled LAB	2	0.0958	0.2479
ResNet 18	RGB	5	0.1227	0.2711
ResNet 18	LAB	5	0.1351	0.2945
ResNet 18	Re-scaled LAB	5	0.1404	0.3015
SimpleNet 18	RGB	5	0.1081	0.2476
SimpleNet 18	LAB	5	0.1093	0.2558
SimpleNet 18	Re-scaled LAB	5	0.1266	0.2799
DAE S&P 5% Noise	RGB	5	0.0653	0.1815

Table 1: Summary of Experiments

- 500k Unlabelled Images (NYU Dataset)
- Image Size: 96x96
- 5-10 Epochs of Pre-training
- Best Image Space: Re-Scaled LAB
- Best Model: ResNet 18

Up to 14.05% Top-1 Accuracy on NYU Data Classification

References

[1] Richard Zhang, Phillip Isola, and Alexei A. Efros. Split-brain autoencoders: Unsupervised learning by cross-channel prediction. *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Jul 2017.

[2] G.E. Hinton and R.R. Salakhutdinov. Reducing the dimensionality of data with neural networks. *Science (New York, N.Y.)*, 313:504–7, 08 2006.

[3] Pascal Vincent, Hugo Larochelle, Yoshua Bengio, and Pierre-Antoine Manzagol. Extracting and composing robust features with denoising autoencoders. In *Proceedings of the 25th International Conference on Machine Learning*, ICML '08, pages 1096–1103, New York, NY, USA, 2008. ACM.

[4] Pascal Vincent, Hugo Larochelle, Isabelle Lajoie, Yoshua Bengio, and Pierre-Antoine Manzagol. Stacked denoising autoencoders: Learning useful representations in a deep network with a local denoising criterion. *J. Mach. Learn. Res.*, 11:3371–3408, December 2010.

[5] Jonathan Masci, Ueli Meier, Dan C. Ciresan, and Jürgen Schmidhuber. Stacked convolutional auto-encoders for hierarchical feature extraction. In *ICANN*, 2011.

[6] Michele Alberti, Mathias Seuret, Rolf Ingold, and Marcus Liwicki. A pitfall of unsupervised pre-training, 2017.

[7] Mehdi Noroozi and Paolo Favaro. Unsupervised learning of visual representations by solving jigsaw puzzles. *CoRR*, abs/1603.09246, 2016.

[8] Alexey Dosovitskiy, Jost Tobias Springenberg, Martin A. Riedmiller, and Thomas Brox. Discriminative unsupervised feature learning with convolutional neural networks. *CoRR*, abs/1406.6909, 2014.

[9] Phillip Isola, Daniel Zoran, Dilip Krishnan, and Edward H. Adelson. Learning visual groups from co-occurrences in space and time. *CoRR*, abs/1511.06811, 2015.

[10] Alireza Makhzani and Brendan J. Frey. A winner-take-all method for training sparse convolutional autoencoders. *CoRR*, abs/1409.2752, 2014.

[11] Junbo Jake Zhao, Michaël Mathieu, Ross Goroshin, and Yann LeCun. Stacked what-where auto-encoders. *CoRR*, abs/1506.02351, 2015.

[12] Shixiang Gu and Luca Rigazio. Towards deep neural network architectures robust to adversarial examples. 12 2014.

[13] Junyuan Xie, Linli Xu, and Enhong Chen. Image denoising and inpainting with deep neural networks. In F. Pereira, C. J. C. Burges, L. Bottou, and K. Q. Weinberger, editors, *Advances in Neural Information Processing Systems 25*, pages 341–349. Curran Associates, Inc., 2012.

[14] Deepak Pathak, Philipp Krähenbühl, Jeff Donahue, Trevor Darrell, and Alexei A. Efros. Context encoders: Feature learning by inpainting. *CoRR*, abs/1604.07379, 2016.

[15] Terrance Devries and Graham W. Taylor. Improved regularization of convolutional neural networks with cutout. *CoRR*, abs/1708.04552, 2017.

[16] Richard Zhang, Phillip Isola, and Alexei A. Efros. Colorful image colorization. *CoRR*, abs/1603.08511, 2016.

[17] Gustav Larsson, Michael Maire, and Gregory Shakhnarovich. Colorization as a proxy task for visual understanding. *CoRR*, abs/1703.04044, 2017.

[18] Gustav Larsson, Michael Maire, and Gregory Shakhnarovich. Learning representations for automatic colorization. *CoRR*, abs/1603.06668, 2016.

[19] J. Long, E. Shelhamer, and T. Darrell. Fully convolutional networks for semantic segmentation. In *2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 3431–3440. June 2015.

[20] Alex Krizhevsky, Ilya Sutskever, and Geoffrey E Hinton. Imagenet classification with deep convolutional neural networks. In F. Pereira, C. J. C. Burges, L. Bottou, and K. Q. Weinberger, editors, *Advances in Neural Information Processing Systems 25*, pages 1097–1105. Curran Associates, Inc., 2012.

[21] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. *CoRR*, abs/1512.03385, 2015.

[22] Seyyed Hossein HasanPour, Mohammad Rouhani, Mohsen Fayyaz, and Mohammad Sabokrou. Lets keep it simple, using simple architectures to outperform deeper and more complex architectures. *CoRR*, abs/1608.06037, 2016.