社団法人　電子情報通信学会
THE INSTITUTE OF ELECTRONICS,
INFORMATION AND COMMUNICATION ENGINEERS

信学技報
TECHNICAL REPORT OF IEICE.

PyramidNet における確率的な正則化の効果の検証

山田　良博†　岩村　雅一†　黄瀬　浩一†

† 大阪府立大学大学院工学研究科
〒 599–8531 大阪府堺市中区学園町 1–1
E-mail: yamada@m.cs.osakafu-u.ac.jp, {masa,kise}@cs.osakafu-u.ac.jp

あらまし　画像を『山』『椅子』『ラーメン』といったカテゴリで分類する問題を一般物体認識と呼ぶ．しかし高精度な一般物体認識は膨大な枚数の画像での学習を行うことが一般的であり，データセットの作成コストや学習時間の観点から現実的でない．そこで限られた枚数でも高い精度を実現する一般物体認識手法が求められている．限られたデータ量でも高い精度を実現する一般物体認識手法の 1 つが PyramidNet である．PyramidNet は従来，幾つかの段階で行われていた処理を逐次的に行うことによって，従来手法から飛躍的に認識精度を高めた手法である．しかし PyramidNet は従来手法で精度を高めるために用いられていた確率的な正則化処理を含んでおらず，性能が頭打ちになっている可能性がある．本稿では PyramidNet に適した確率的な正則化手法を検討するべく評価実験を行った．

キーワード　一般物体認識，深層学習，Deep Residual Network，Residual Learning，正則化

1. はじめに

一般物体認識において，同じカテゴリに属する様々な「見え」を持つ物体の認識は中心的な課題である．この実現には，物体の色や形状などの見えの変化を吸収し，本質的な特徴を見出す必要がある．2012 年以降，Convolutional Neural Network (CNN) が本質的な特徴を見出し優れた性能を発揮すると注目を浴びている [1]．CNN は，画像の畳み込みを実現する複数の畳み込み層から成り，畳み込み層が増える程，抽象化した特徴を取り出すことができる．そのため，総数が増えれば，より複雑な特徴を持つカテゴリが認識できるようになると考えられている．しかし，これは諸刃の剣であり，多くの畳み込み層を持つ深層ネットワークで単純な特徴で構成されるカテゴリをうまく表現できず，認識精度が頭打ちになることが報告されている [2]．これは，ネットワークの前段の少数の畳み込みで得られた，比較的単純な特徴が後段の畳み込みによって潰れてしまうからと考えられている．

この従来の CNN の問題を解決したのが ResNet [3] である．ResNet の最大の特徴は，図 1 に示す Residual Unit の導入である．Residual Unit は，従来のように畳み込みを行う場合と，この層への入力をそのまま出力して畳み込みを行わない場合の結果を足し合わせる処理機構によって，畳込みが非常に多い CNN ながら，特徴を潰さず高い精度を実現している．ResNet は図 2 のような Residual Unit を多数含む構造になっている．ResNet の登場以降，Residual Unit の構造をいかに改善するかが CNN における大きな課題となっている．

ResDrop [4] は ResNet の改良手法の一つである．ResNet は従来より多くの畳み込みを持つため，Residual Unit の導入をもってしてもなお，得られた特徴が学習の途中で失われてしまう効果が無視できない．また，学習時間が長いという問題もある．この問題の解決方法として，ResDrop では，毎回ランダム

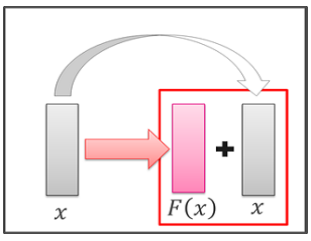

図 1　ResNet の Residual Unit の基本構造．入力 x に対して畳み込みを行う場合の出力 $F_{(x)}$ と畳み込みを行わない場合の出力 x を足し合わせる構造をもつ．

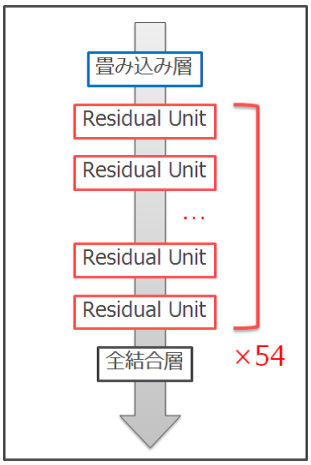

図 2　ResNet の全体図．Residual Unit は 1 つあたり 2 つの畳み込みを含むため，Residual Unit を 54 個持つこの ResNet は，畳み込みを行う処理層と全結合と呼ばれる処理を行う処理層を全て合わせて 110 層の処理層で構成されている．

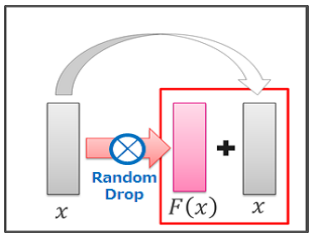

図 3　ResDrop の Residual Unit の基本構造．ResNet の構造に対して，確率的に畳み込みの出力を 0 にする Random Drop を導入する．

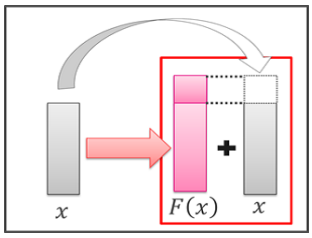

図 4　PyramidNet の Residual Unit の基本構造．ResNet の構造に対して，channel の次元数を増加させる．

に選ばれた一部の畳み込みを学習時に無視する Random Drop と呼ばれる確率的な正則化が導入された．ResDrop の基本構造を 3 に示す．これにより，学習時には畳み込み数の少ない CNN のように扱うことができ，認識時には ResNet 本来の性能を発揮することができ，認識精度の向上と学習時間の削減が報告されている．

PyramidNet [5] は ResDrop とは異なる ResNet の改良手法である．ResNet では，いくつかの Residual Unit で channel と呼ばれる特徴の次元数を急激に増加させている．この channel が急激に増加する Residual Unit は他の Residual Unit とは異なる傾向をもつ学習を行っており，認識精度を低下させる一因となり得ることが指摘されている [6]．この問題の解決方法として，PyramidNet では，channel が急激に増加する Residual Unit を失くし，代わりに図 5 のように各 Residual Unit で緩やかに channel を増加させる機構が導入された．これにより，認識精度の向上が報告されている．

ResDrop と PyramidNet は互いに異なる工夫で ResNet の改良を実現している．表 1 にまとめたように ResDrop は channel が急激に増加する Residual Unit への対策を持たず，PyramidNet は非常に多くの畳み込みを持つことへの対策を持たない．そこで両者の弱点を補い合うことで図 5 に示すように新たな ResNet の改良手法が実現できると考えられる．このように両者を単純に組み合わせるアイディアは PyramidNet の論文中で言及はされているものの、具体的な実験結果は報告されておらず，PyramidNet と ResDrop の組み合わせがどのような効果をもたらすかについて充分に検討されていない．本稿では図 5 の手法を PyramidDrop と呼ぶことにする．本稿ではさらに，図 6 に示す PyramidDrop の改良手法を提案する．これを PyramidSepDrop と呼ぶことにする．我々は PyramidNet に ResDrop の正則化を適用する効果を実験的に検討した．そ

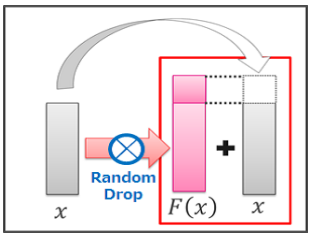

図 5　PyramidDrop ([5] ならびに本稿) の基本構造．PyramidNet の出力に Random Drop を導入する．

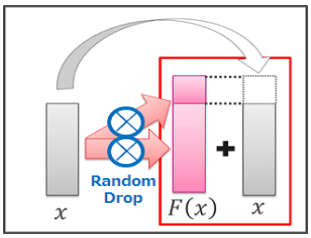

図 6　提案手法 PyramidSepDrop の基本構造．PyramidNet における channel の増加部分とそうでない部分では本質的に異なる学習を行っていると考え，Residual Unit の畳み込みの出力を分岐させ，独立に Random Drop を導入する．

してます，PyramidNet と ResDrop を組み合わせた場合でも ResDrop の確率的な正則化が有効になることを確認した．さらに，PyramidDrop と PyramidSepDrop は表 1 に示す他の手法とは異なり，通常は処理の高速化のために行われるデータパラレルな並列学習が認識精度の向上にも寄与するという興味深い結果が得られた．これらの検討の中で提案手法は一般物体認識用データセット CIFAR-10 および CIFAR-100 を用いた実験で，表 1 に示すように原稿執筆時点 (2017 年 1 月 29 日) においての世界最高の認識精度を達成している．

2. 先 行 研 究

本章では，ResNet，ResDrop，PyramidNet について説明する．

2.1 ResNet

前述のように CNN は畳み込み層が多いと様々な特徴を取り出すことができるが，得られた重要な特徴が多くの畳み込み層の中で消えてしまうことがある．これは畳み込みで恒等写像 $f(x) = x$ を実現することが難しいことに起因する．恒等写像を学習することができれば，得られた重要な特徴が畳み込み層の中で消えてしまいにくくなると考えられる．

そこで恒等写像を扱うため，図 1 に示す Residual Unit が提案された．これは畳み込みによる写像と入力を足し合わせて出力とするもので，入力を x としたとき，以下のように表される．ただし $G(x)$ は入力 x に対する Residual Unit 全体の変換であり，$F(x)$ は Residual Unit の畳み込み部分のみの変換である．

$$G(x) = x + F(x) \tag{1}$$

Residual Unit では恒等写像を扱うとき，常に $F(x) = 0$ になる畳み込みを学習することになる．これは $f(x) = x$ よりも簡単に実現できるため，従来の畳み込みに比べて恒等写像の学習

表 1　各手法とその特徴および一般物体認識用データセット CIFAR-10 と CIFAR-100 を用いた際のエラー率．ResDrop は Residual Unit に確率的な正則化を導入することで精度を向上させている．PyramidNet は channel の増加を幾つかの Residual Unit で行うのではなく，各 Residual Unit で行うことで精度を大幅に向上させている．提案手法は両方の工夫を組み合わせることで精度向上を達成した．

手法	正則化	channel	CIFAR-10	CIFAR-100
ResNet [3]	×	×	6.43%	25.16%
ResDrop [4]	○	×	5.23%	24.58%
PyramidNet [5]	×	○	3.77%	18.29%
PyramidDrop ([5] ならびに本稿)	○	○	-	**16.28%**
PyramidSepDrop (本稿)	○	○	**3.31%**	**16.18%**

が容易になると考えられる．実際に Residual Unit を大量に積み重ねた ResNet は従来の CNN に比べて大きく精度が改善しており，ImageNet [7] を用いた実験では人の平均的な認識精度を超えるまでになっている．

2.2 ResDrop

ResDrop は式 (1) の畳み込み部分 $F(x)$ について，確率的に $F(x) = 0$ とする処理機構を導入し，毎回の学習過程でランダムに学習を行わない層を決定する．この工夫によってそれぞれの Residual Unit が特徴を補い合い精度の高い特徴を取り出せるようになり，認識精度を高めながら学習時間を削減できる．ResDrop を導入することで ResNet に比べ精度が高くなること，特に ResNet で最も認識精度が高かった 110 層の処理層を持つ CNN から Residual Unit を大幅に増やした 1202 層の学習で精度が向上することが確認されている．

2.3 PyramidNet

図 4 に示す Residual Unit を持つ PyramidNet は ResDrop と同様に ResNet の認識精度を向上させる手法である．CNN は特徴を抽出する過程で，画像が持つ高さや幅とは異なるもう一つの次元，channel に対して処理を行う．ResNet ではいくつかの Residual Unit でこの channel に関する出力が大きく増加する．この channel が急激に増加する Residual Unit では，認識精度に直結する特徴が取り出されることが実験的に確認されている．ResDrop はその働きを軽減させ認識精度を向上させているため，channel が急激に増加する Residual Unit が認識精度の向上を妨げていると考えられる [6]．したがって channel の次元数は大きく変化しないことが望ましい．ただし channel 数が少なければ充分な特徴を抽出できず，多すぎるとメモリ容量が足りなくなるため，入力付近の channel は少なく出力付近の channel が多くなるよう調整する必要がある．そこで PyramidNet はいくつかの Residual Unit で channel を急激に増加させるのではなく，各 Residual Unit で channel を徐々に増加させることで，急激な channel の増加に関する問題の解決を図っている．PyramidNet はこの工夫によって，一般物体認識データセット CIFAR-10 および CIFAR-100 において高い認識精度を実現している．

3. 提 案 手 法

本章では，具体的な提案手法について検討する．

3.1 PyramidDrop

図 5 に示す PyramidDrop は PyramidNet と ResDrop を組み合わせることで更なる認識精度の向上を目指した手法である．PyramidDrop は PyramidNet の出力に対して一様に Random Drop を導入している．PyramidDrop は PyramidNet の論文中で言及されているものの、具体的な実験結果は報告されていない．ただし 2 章で述べたように ResDrop は多層な CNN で精度を改善していることから，多層な PyramidNet での認識精度の向上が期待できる．

3.2 PyramidSepDrop

図 6 に示す PyramidSepDrop は PyramidDrop を更に改善することを目指した手法である．1 章で述べたように，ResNet には幾つかの急激に channel が増加する Residual Unit が含まれており，channel が急激に増加する Residual Unit は他の Residual Unit とは異なる傾向を持つ学習を行うため，認識精度を低下させる一因となり得ることが指摘されている．ここで channel が増加する Residual Unit における $F(x)$ は，前層の出力を受け取る channel と新たに増えた channel の 2 つの部分から成り，それぞれ性質が異なると考えられる．前者は前層までに得られた特徴の強化を図り，後者は新たな特徴を畳み込みで生成する．そのため，これらは独立に考える必要がある．PyramidNet では緩やかに channel を増加させることで，急激に channel が増加する Residual Unit の問題の解決を図っているが，channel の増加は行われているため，同様の問題は依然として存在する．そこで，channel の増加を考慮した確率的な正則化の導入が有効であると考える．

PyramidNet で用いられる畳み込み等の処理は channel 毎に独立であるため，これは大きな問題にならない．しかし Random Drop は channel を独立に扱わない処理であるため，分離して考える必要がある．PyramidSepDrop は Residual Unit の中で畳み込みの出力を分岐させ，channel 増加部分とそうでない部分についてそれぞれ独立な Random Drop を導入している．この工夫によって PyramidDrop と同様に学習時には畳み込みの少ない CNN のように扱うことができる．また channel 増加部分やそうでない部分だけ出力した場合を考慮できるため，PyramidDrop よりも柔軟なネットワーク構造で学習を進めることが出来る．この特徴から PyramidDrop よりも比較的層が浅いネットワークでも性能を改善することが期待できる．

4. 実　　験

それぞれの手法について比較を行い，提案手法の有効性を検証した．実験 1 ではそれぞれの手法について，データパラレルな並列化により，認識精度にどのような影響が出るのかを確認した．実験 2 では PyramidNet と提案手法について，層数を増減させた場合や PyramidNet のパラメータの一つである channel の増加量を変化させた場合に，認識精度にどのような影響が出るのかを確認した．実験 3 では実験 1，2 の結果を踏まえて，より高い精度で画像が可能か検討した．

4.1 実　験　1

ResNet，ResDrop，PyramidNet，提案手法を用いて，並列に学習するモデル数を増やした際の認識精度を確かめた．

データセットは CIFAR-10 および CIFAR-100 を用いた．ResDrop，及び提案手法の Random Drop における死亡率は，最初の Residual Unit から一定で増加し最後の Residual Unit で 0.5 となるよう設定した．ResNet 及び ResDrop については以下の条件に従った．層数は 110，Epoch は 163，BatchSize は 128，重み減衰は 0.0001，モメンタムは 0.9，Nesterov の加速法を用い，初期学習率を 0.1 とした．学習率は Epoch が 81 の時点で 0.01，122 の時点で 0.001 となるように設定した．モデル数は 1，4，8，16 のいずれかを用いた．ResDrop は確率的な正則化を導入した以外は ResNet と同じ構造を利用した．PyramidNet 及び提案手法については以下の条件に従った．層数は 110，Residual Unit あたりの channel の増加は 5，Epoch は 300，BatchSize は 128，重み減衰は 0.0001，モメンタムは 0.9，Nesterov の加速法を用い，初期学習率を 0.5 とした．学習率は Epoch が半分進んだ時点で 0.05，4 分の 3 進んだ時点で 0.005 となるように設定した．

それぞれの手法の結果を表 2，図 7 および図 8 に示す．ResNet ではモデル数が 8 になるまで並列学習が有効だった．一方，ResDrop はいずれも ResNet よりも並列学習が有効でないことがわかった．PyramidNet では ResNet と同様にモデル数が 8 になるまで並列学習が有効だった．一方，ResDrop では並列学習が有効でなかったが，提案手法ではいずれもモデル数が 16 になるまで並列学習が有効であった．また，PyramidSepDrop は PyramidDrop よりもモデル数が少なくても高い精度を実現することがわかった．

表 2　層数 110 の場合の各手法における最終 Epoch のエラー率．

手法	モデル数	CIFAR-10	CIFAR-100
ResNet	1	-	30.54%
	4	-	28.58%
	8	-	27.16%
	16	-	27.56%
ResDrop	1	-	26.56%
	4	-	25.63%
	8	-	26.09%
	16	-	25.92%
PyramidNet	1	3.77%	18.29%
	4	-	17.87%
	8	-	17.80%
	16	-	17.97%
PyramidDrop	1	3.99%	18.30%
	4	-	17.78%
	8	-	17.40%
	16	-	17.05%
PyramidSepDrop	1	3.66%	18.01%
	4	-	17.53%
	8	-	17.28%
	16	-	17.12%

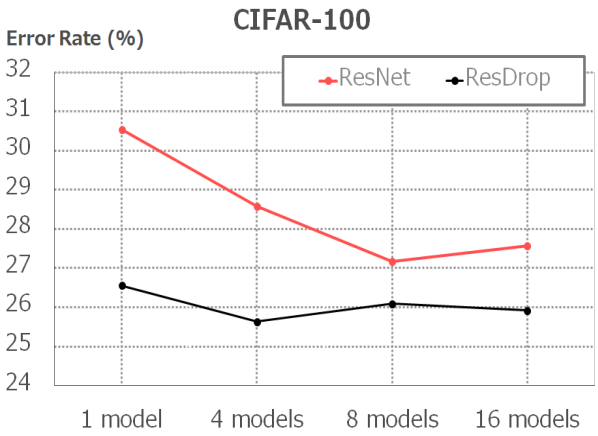

図 7　層数 110 の場合の ResNet および ResDrop における最終 Epoch のエラー率．

表 3　層数 56，Residual Unit あたりの channel の増加が 5，モデル数が 4 の場合の PyramidNet および提案手法における最終 Epoch のエラー率．

手法	層数	channel の増加	CIFAR-10	CIFAR-100
PyramidNet	56	5	-	20.27%
PyramidDrop	56	5	-	20.36%
PyramidSepDrop	56	5	-	20.22%

4.2 実　験　2

実験 1 と同様の条件で，モデル数を 4 とした場合の PyramidNet および提案手法について層数や増加する channel の数を変化させた際の認識精度を確かめた．

層数が 56，Residual Unit あたりの channel の増加が 5 の場合の結果を表 3 に示す．いずれについても大きな変化はなかった．

層数が 56，Residual Unit あたりの channel の増加が 10 の場合の結果を表 4 に示す．PyramidNet や PyramidDrop に比べ PyramidSepDrop の精度が高くなった．

層数が 182，Residual Unit あたりの channel の増加が 5 の場合の結果を表 5 に示す．PyramidNet のエラー率が比較的高く，PyramidDrop と PyramidSepDrop のエラー率が並ぶ結果となった．

4.3 実　験　3

実験 1，2 と同様の条件で，182 層，channel の増加が 5 でモデル数が 4，16 の場合の PyramidSepDrop について，CIFAR-10 および CIFAR-100 における認識精度を確かめた．

モデル数が 4 の場合と 16 の場合の比較を表 6 に示す．モデ

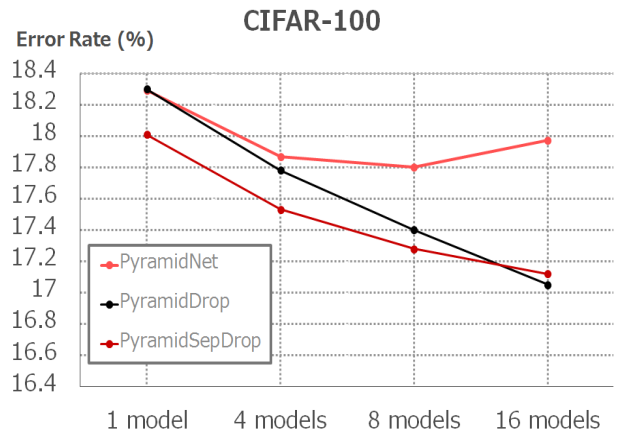

図 8　層数 110 の場合の PyramidNet および提案手法における最終 Epoch のエラー率.

表 4　層数 56，Residual Unit あたりの channel の増加が 10，モデル数が 4 の場合の PyramidNet および提案手法における最終 Epoch のエラー率.

手法	層数	channel の増加	CIFAR-10	CIFAR-100
PyramidNet	56	10	-	19.33%
PyramidDrop	56	10	-	18.64%
PyramidSepDrop	56	10	-	18.19%

表 5　層数 182，Residual Unit あたりの channel の増加が 5，モデル数が 4 の場合の PyramidNet および提案手法における最終 Epoch のエラー率.

手法	層数	channel の増加	CIFAR-10	CIFAR-100
PyramidNet	182	5	-	17.13%
PyramidDrop	182	5	-	16.28%
PyramidSepDrop	182	5	3.45%	16.33%

表 6　層数 182，Residual Unit あたりの channel の増加が 5，モデル数が 4 および 16 の場合の提案手法 PyramidSepDrop における最終 Epoch のエラー率.

手法	層数	モデル数	CIFAR-10	CIFAR-100
PyramidSepDrop	182	4	3.45%	16.33%
	182	16	3.31%	16.18%

ル数を増やすことで僅かながら認識精度が向上することを確認した.

4.4　考　　察

これらの結果から，PyramidNet における確率的な正則化手法はモデル数の増加に応じて認識精度を向上させる効果があると考えられる.

実験 1 において ResNet と PyramidNet では並列化に応じて同様の推移が見られたが，ResDrop と提案手法では異なる推移が見られた．また，実験 2 において層数や channel の増加に応じて PyramidNet と提案手法の精度が変化することが確認された．これらのことから提案手法における並列化に応じた精度改善は PyramiNet の特殊な channel の増加法に起因すると考えられる.

表 7　主な CNN と提案手法 PyramidSepDrop のエラー率.

手法	CIFAR-10	CIFAR-100
ResNet	6.43%	25.16%
ResDrop	5.23%	24.58%
DenseNet [9]	3.74%	19.25%
PyramidNet	3.77%	18.29%
ResNeXt [8]	3.58%	17.31%
DenseNet-BC [9]	-	17.18%
PyramidSepDrop	**3.31%**	**16.18%**

今回の実験条件では 2 種類の提案手法に関してはいずれの結果においても PyramidSepDrop が PyramidDrop と同等，もしくは上回ることが確認された．特にモデル数が小さい場合に PyramidSepDrop が優れている傾向が確認されたことから，PyramidSepDrop は並列化と同様の効果を与えている可能性が高い．ただし層数やモデル数が多い大規模な条件では非常に僅かながら PyramidDrop が PyramidSepDrop の認識精度を上回る傾向が見られたため，実験 3 のような条件や，より大規模な条件で確認する必要があると考えられる.

5.　まとめと今後の課題

本稿では ResDrop と PyramidNet を組み合わせた手法 PyramidDrop 及び PyramidSepDrop を提案し，その効果を実験的に検討した．その結果，表 7 に示すように，CIFAR-100 において，提案手法のベースになった PyramidNet からは 2.11%，本稿の検討には含まれていないが，ResNeXt [8] からは 1.13%、DenseNet-BC [9] からは 1.00%など，従来手法から認識精度が大幅に改善された．今後の目標として，CIFAR-10 の検証を進めること，大規模な条件に関する検証を進めること，更なるエラー率の低下を目指したパラメータ検証を行うことを目標とすることが挙げられる.

謝辞　本研究は，JST CREST，JSPS 科研費 25240028 ならびに AWS Cloud Credits for Research program の補助による.

文　　献

[1] A. Krizhevsky, I. Sutskever, and G.E. Hinton, “Imagenet classification with deep convolutional neural networks,” Advances in Neural Information Processing Systems 25, 2012.
[2] K. He and J. Sun, “Convolutional neural networks at constrained time cost,” CoRR, vol.abs/1412.1710, , 2014. http://arxiv.org/abs/1412.1710
[3] K. He, X. Zhang, S. Ren, and J. Sun, “Deep residual learning for image recognition,” Proc. CVPR, 2016.
[4] G. Huang, Y. Sun, Z. Liu, D. Sedra, and K. Weinberger, “Deep networks with stochastic depth,” CoRR, 2016. v3. https://arxiv.org/abs/1603.09382
[5] D. Han, J. Kim, and J. Kim, “Deep pyramidal residual networks,” CoRR, 2016. https://arxiv.org/abs/1610.02915
[6] A. Veit, M.J. Wilber, and S. Belongie, “Residual networks behave like ensembles of relatively shallow networks,” Advances in Neural Information Processing Systems 29, 2016.
[7] J. Deng, W. Dong, R. Socher, L.-J. Li, K. Li, and L. Fei-Fei, “ImageNet: A Large-Scale Hierarchical Image Database,” CVPR09, 2009.
[8] S. Xie, R.B. Girshick, P. Dollár, Z. Tu, and K. He, “Aggregated residual transformations for deep neural networks,”

2016. http://arxiv.org/abs/1611.05431

[9] G. Huang, Z. Liu, and K.Q. Weinberger, "Densely connected convolutional networks," CoRR, 2016. https://arxiv.org/abs/1608.06993