

画像分類精度を向上させる追加タスクの検討

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 画像中に含まれる物体を分類する画像分類技術において、Convolutional Neural Network (CNN) は分類精度を向上させることで注目を集めている。原則として、CNN の学習は各画像に正しい分類結果を表すラベルが付与されているという条件が仮定されている。この場合、優れた特徴表現能力を学習するためには、人手でラベルが付けられた膨大な枚数の画像を学習する必要がある、コストが高い。この状況を改善するために、ラベルが付与されていない条件で優れた特徴表現能力を獲得する手法が模索されていた。近年、CNN に対して本来の画像分類とは異なる学習タスク (Pretext Task) を扱うことで、ラベルが付与されていない条件で優れた特徴表現が学習可能なことが示された。このような Pretext Task は従来の教師あり学習とは異なる特徴表現能力を学習している可能性が高く、組み合わせることで分類精度を改善できると考えられる。しかし、Pretext Task と本来の画像分類タスクを組み合わせた場合については詳しく検証されておらず、有効な方法は示されていない。そこで、2 種類のアプローチを通して、Pretext Task と本来の画像分類タスクについての有効性について検討を行った。実験結果からは、全くランダムに与えられたラベルの学習であっても、本来の画像分類タスクに寄与することを確認した。ただし、我々が検討したタスクは認識精度を大きく改善しなかった。

キーワード 一般物体認識, 深層学習, Pretext Task

1. はじめに

画像中のオブジェクトを分類する画像分類技術は画像処理の根幹を成す重要な技術の 1 つであり、画像等を用いた応用技術を実現する上で、画像分類精度の改善は欠かすことの出来ない重要な課題である。画像分類技術の中でも、Convolutional Neural Network (CNN) は、大規模な画像データから上手く抽象的な特徴を表現し、優れた分類性能を発揮することで注目を浴びている [1]。しかし、原則として、ニューラルネットワークの学習は教師あり学習 (Supervised Learning) と呼ばれる、各画像に正しい分類結果を表すラベルが付与されているという条件が仮定されている。この場合、優れた特徴表現能力を学習するためには、人手でラベルが付けられた膨大な枚数の画像を学習する必要がある、コストが高い。この状況を改善するために、別のデータで学習したニューラルネットワークを転用する転移学習 (Transfer Learning) [2]、ラベルの一部のみが付与された場合を考える弱教師あり学習 (Semi-Supervised Learning) [3] やラベルが付与されていないデータを用いる教師なし学習 (Unsupervised Learning) [4]、教師なし学習の問題設定において自明なクラスラベルが付与できるタスクを用意して、ニューラルネットワークの特徴表現能力をかさ上げる自己教師あり学習 (Self-Supervised Learning) [5] 等の技術が注目を集めている。

前述の自己教師あり学習の中で、本来の画像分類とは異なる

学習タスク (Pretext Task) の学習によって、本来の教師あり学習を用いた画像分類の分類精度 80% に対して、ラベルが付与されていないにも関わらず 7% から 10% 程度しか劣らない優れた特徴の抽出が学習可能であることが示された [5-7]。これらの報告は、Pretext Task による教師なし学習で得られる特徴表現能力が画像分類タスクに大きく寄与することを示している。

ここで疑問となるのは、どのような Pretext Task が画像分類タスクの特徴表現能力の獲得に寄与するのかである。画像の回転 [5] や、バラバラに分割され並び変えられた画像の正しい並びを推定する [6, 8] といった Pretext Task の有効性が既に報告されている。これらは画像分類の学習とは異なる対象を Pretext Task として学習しており、異なる特徴表現能力を獲得すると考えられる。そこで、画像分類と同時に有用な Pretext Task を学習することが出来れば、異なる特徴表現能力の組み合わせを学習でき、画像分類精度を向上させることが出来ると考えられる。

しかし、これらの Pretext Task を用いた画像分類精度の向上には二つ疑問が存在する。一つは、Pretext Task と本来の画像分類タスクとの関連が明らかでないため、具体的にどのようなタスクが特徴表現能力の獲得に寄与しているのかわからない点である。より有効な組み合わせを発見するためには、タスクと特徴表現能力の獲得の分析は不可欠と考えられる。従って、画像分類タスクとの関連が明らかでないタスクを用いることで、特徴表現能力とは何かを分析する必要がある。もう一つは、画像

分類タスクと同時に利用した際に学習を阻害する点である。画像に対する大きな回転や分割と並び替えは、本来の画像の性質を損ね分類に対して強い影響を与えるため、画像分類精度に悪影響を及ぼしうる。従って、画像分類精度に悪影響を及ぼさない画像処理を行う Pretext Task を検討する必要がある。

そこで我々は、二つの疑問を一挙に解決しうるものとして従来の画像分類システムの枠組みに則る新たな追加タスクを検討した。我々が検討したのは大きく分けて二種類の追加タスクである。

- ラベルを推定する追加タスク
- Data Augmentation を当てる追加タスク

前者は本来の画像分類タスクと同様の枠組みで異なるラベルを推定することを追加タスクとしたものである。異なるラベルであることを除けば本来の画像分類タスクそのものであり、関連が明らかである。また追加の画像処理を行わない。後者は本来の画像分類タスクで広く用いられるデータ水増し法のパラメータを推定することを追加タスクとしたものであり、前者と同様画像分類タスクの枠組みになり、追加の画像処理を行わない。これらは二つの疑問に対する条件を満たすものであり、画像分類精度を向上させる追加タスクとして妥当なものである。

実際にこれらの追加タスクについて画像データセット CIFAR-100 に対する影響を分析した。その結果、全くランダムに与えられたラベルの学習であっても、本来の画像分類タスクに寄与することを確認した。ただし、我々が検討した追加タスクは分類精度を僅かに改善したものの、大きく改善しなかったが、興味深い結果が得られたため、情報共有できればと考えている。

2. 関連研究

特徴表現に関しては、歴史的な背景から CNN が実験的な形で有効性が示されてきたために、CNN がいかにして特徴表現を獲得するのか基礎的解析が進んでいなかった。しかし近年は幾つかの重要な発見に伴い基礎的解析が進展しており、今後は特徴表現について関連が深い研究として基礎的解析と Pretext Task の分析が融合していくと予想される。そこで本節では、我々の研究内容に関連が深い一連の研究として、CNN における重要な特徴表現に対する基礎的解析の進展について述べる。また、現時点で有効な特徴表現が得られることが示されている主要な Pretext Task について説明する。

2.1 特徴表現に対する基礎的解析の進展

近年の研究では、CNN はパラメータ数が増えるにつれて分類精度が向上することが示されている。パラメータ数については「パラメータ数が十分あれば、ランダムにパラメータが初期化された際に、学習に寄与するパラメータを偶然含む」とする宝くじ券仮説が提唱されており、パラメータ数と分類精度の関係性が実験的に示されている [9]。一方、パラメータ数が多い近年の CNN は学習対象にランダムなラベルを与えられた際さえ、容易に適合してしまうことが報告されている [10]。これらの研究結果は、実際には CNN が学習に応じて特徴表現能力を会得していることを示しており、ランダムなラベルさえ推定し得る柔軟な特徴表現能力が獲得可能なことを示唆している。

しかし、どのような特徴表現能力が認識精度に寄与するかは示されていない。

特徴表現能力と認識精度の関係については、ImageNet を用いて教師あり学習された CNN の特徴表現能力について、オブジェクトのテキストチャを重視しておりオブジェクトの形状にはあまり影響を受けていないという傾向が昨年報告されている [11]。これはオブジェクトの形状を重視する人間の物体認識と乖離があり、頑健さや分類精度を損ねている可能性がある。一方で、物体認識における頑健さと分類精度にはトレードオフの関係があることが指摘されている [12]。現時点で特徴表現能力に対する基礎的解析は発展段階であり、今後更に進んでいくことが予想される。

2.2 主要な Pretext Task

Gidaris らは回転を与えた画像に対して、元画像に対してどれだけ回転したかを 4 クラス分類する Pretext Task によって特徴表現能力を学習する手法を提案し、分類精度を検証した [5]。提案手法 RotNet は 90° ずつ回転する画像を推定するだけでも関わらず、CIFAR-10、ImageNet 等の従来の教師なし学習手法による分類精度を凌駕し、教師あり学習と同等の分類精度を達成した。

Noroozi らはバラバラに分割した画像を並び替えた際に、正しい並びを推定する「ジグゾーパズル」を解く Pretext Task によって特徴表現能力を学習する手法を提案し、分類精度を検証した [6, 8]。提案手法は ImageNet 等の大規模データセットで精度が検証されており、当時の教師なし学習手法において高い分類精度を達成した。

Noroozi らは「ある画像の特徴」と「ある画像を分割して得られた 4 個の分割画像に含まれる特徴の総和」が一致するという制約によって特徴表現能力を学習する手法を提案し、分類精度を検証した [7]。提案手法は ImageNet 等の大規模データセットで高い精度が確認されており、ジグゾーパズルと同等かそれ以上の特徴表現能力を獲得することを示した。

これらの Pretext Task は、画像に対して強い変換が与えられ、その変換をラベルとして学習することに相当する。従って、画像分類タスクと異なる画像と異なるラベルに基づき特徴表現能力を学習しているといえ、Pretext Task によって得られる特徴表現能力と目的となる画像分類タスクによって得られる特徴表現能力との関連性が不明瞭である。特徴表現能力の分析のためには、画像分類タスクとの関連が明らかな Pretext Task であることが望ましい。また、同時に利用する際には従来の学習を損ねない変換であることが望ましい。

3. 提案手法

3.1 ラベルを推定する追加タスク

追加タスクとして画像分類タスクを用いた場合、ラベルの違いのみに応じて異なる特徴表現能力を学習する。このとき目的の画像分類タスクと追加タスクの関連は、対象とするラベルの関連性のみで説明される。従って、関連性の高いラベルや関連性の低いラベルを用いることで、目的の画像分類タスクと追加タスクの関連性が決定されるため、追加タスクによって獲得さ

れる特徴表現能力の影響を容易に分析可能となる。

我々は画像データセット CIFAR-100 に対して関連性の異なる 3 種類のラベルセットを準備した。

- Ground Truth である 100 種のラベル
- Ground Truth に対応する 20 種のスーパークラスラベル
- ランダムな 100 種のラベル

3 種類のラベルのうち Ground Truth である 100 種のラベルは学習対象そのものであるため最も関連が高く、ランダムな 100 種のラベルは最も関連が低い。なお、ランダムなラベルの学習を取り扱った先行研究 [10] と同様に、今回のラベルセットの 1 つであるランダムな 100 種のラベルは完全に学習可能であることを確認している。

3.2 Data Augmentation を推定する追加タスク

画像分類タスクの学習の際には、データを擬似的に水増しすることで分類精度を向上させるテクニックとして、小さな摂動を与える Data Augmentation が利用される。そこで、追加タスクとして Data Augmentation に用いたパラメータを同時に推定することで、従来の画像分類タスクの枠組みの中で異なる特徴表現を学習できると考えられる。

我々は画像データセット CIFAR-100 に対して同様に 3 種類の追加タスクを準備した。

- Random Crop + 分類
- Cutout + Semantic Segmentation
- RICAP + Semantic Segmentation

Random Crop は、学習対象よりも大きなサイズにリサイズした画像をランダムな位置で切り出すことで、データを擬似的に水増しする手法であり、Data Augmentation として広く用いられる手法の 1 つである。追加タスクでは、切り出し位置として指定される 9×9 の領域のうち 9 個の 3×3 の格子状の領域を用いて、どの領域に画像の中心が存在するかを分類対象とした。

cutout [13] は、学習画像のランダムな正方形領域をゼロで埋めることで、データを擬似的に水増しする手法であり、近年 Data Augmentation のテクニックとして広く用いられるようになってきている。追加タスクでは、 32×32 の画像中に含まれる最大 16×16 の cutout 領域を推定する Segmentation を学習対象とした。

RICAP [14] は、ランダムに選択された接合点を境に、4 枚の画像を貼り合わせることで新たな画像を作り、データを擬似的に水増しする手法であり、大幅に認識精度を向上させるテクニックの 1 つである。追加タスクでは、 32×32 の各 pixel が接合点を境に左上、右上、左下、右下に存在する 4 枚の画像のうちどの画像に属するかを推定する 4 パターン Segmentation を学習対象とした。

3.3 追加タスクに伴う学習コストの増加

追加タスクの学習に際して、本来の画像分類と同時に学習を行う場合を提案手法の対象とした。この場合の提案手法のネットワーク構造を図 1 に示す。追加タスクの学習に用いられる追加のネットワーク構造にかかるコストが増分となる。同様にネットワーク全体がもつパラメータ数は同時に行う追加タスク

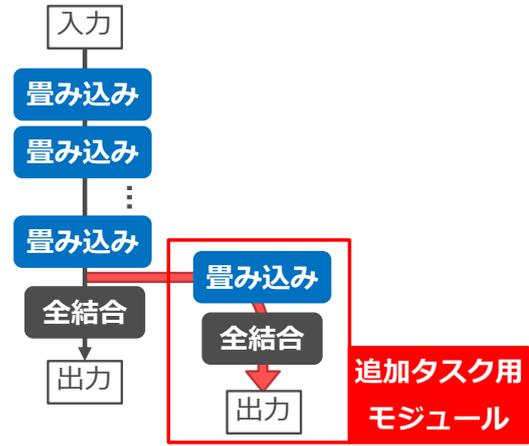


図 1: 提案手法におけるネットワーク構造. 特徴抽出部となる従来のネットワークに対して、新たに追加タスク用モジュールを導入する。追加タスク用モジュールでは追加タスクを対象とした出力を行い、通常の出力と同様に学習する。特徴抽出部には追加タスク用モジュールから勾配が伝播する。追加タスク用モジュールにかかる学習がオーバーヘッドとなるが、従来の画像分類タスクに利用されるパラメータ数に変化はない。

に応じて増加する。ただし、本来の画像分類に直接用いられるパラメータ数については増加していない。

4. 実験

300 Epoch, BatchSize 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 150Epoch 時点で $1/10$, 225Epoch 時点で $1/100$ となるように設定した。GPU を 4 台使用し、並列に学習を行った。全ての画像に対して 3 Channel の平均及び標準偏差を算出し、平均 0 分散 1 になるよう前処理を行った。Data Augmentation として size 4 の padding を施し、 32×32 になるよう Random Crop を行った。更に $1/2$ の確率で Horizontal Flip を行った。Accuracy は最終 Epoch 時点のものを示した。

実験 1 では、ラベルを推定する追加タスクを対象に、3 種類のラベルセットで学習を行い、得られた特徴抽出の有効性を検証した。実験 2 では、Data Augmentation を推定する追加タスクを対象に、3 種類の追加タスクで学習を行った。

4.1 実験 1

3 種類のラベルセットに対する学習を行い、特徴抽出の有効性を評価した。特徴抽出部として PyramidNet [15] の全結合層直前までを用いた構造を利用した。PyramidNet は層数 110 で Basic Block を用い、最終層の Channel が 286 になるもの (PyramidNet-110 $\alpha = 270$) を用いた。

3 種類のラベルセットに対してのみ学習を行った際の Training Accuracy を表 1 に示す。いずれのラベルセットに対しても 100% 近い Training Accuracy となり、適合していることを確認した。

学習した Network に対して、PyramidNet の最終層に位置する全結合層のみを Ground Truth で再学習した場合の結果を

表 1: Ground Truth, スーパークラス, ランダムなラベルセットを学習対象に用いた際の最終 Epoch の Accuracy.

学習ラベルセット	関連性	Network	Train
Ground Truth	高	PyramidNet	99.99%
スーパークラス	中	PyramidNet	99.98%
ランダム	低	PyramidNet	99.90%

表 2: 全結合層のみを Ground Truth で再学習した際の最終 Epoch の Accuracy. 比較のため, 1 層の全結合層のみからなる Network と学習前の初期化された状態の PyramidNet の全結合層のみを学習した結果を併記している.

学習ラベルセット	関連性	Network	Train	Test
Ground Truth	高	PyramidNet	99.99%	80.64%
スーパークラス	中	PyramidNet	77.20%	68.21%
ランダム	低	PyramidNet	28.52%	22.77%
Ground Truth	高	全結合のみ	16.89%	14.53%
Ground Truth	高	PyramidNet (学習前)	1.25%	3.55%

表 3: それぞれの学習で得られたモデルを用いて, Ground Truth に対して転移学習した際の最終 Epoch の Accuracy. ただし比較のため表 3 の結果を併記している.

学習モデル	関連性	Network	Train	Test
Ground Truth	高	PyramidNet	99.98%	81.47%
スーパークラス	中	PyramidNet	99.97%	81.36%
ランダム	低	PyramidNet	99.98%	77.83%
(ランダム初期値)	-	PyramidNet	99.99%	80.64%

表 2 に示す. ただし Ground Truth + PyramidNet は再学習ではなく, 通常学習時の値を記載している. 比較対象として, PyramidNet ではなく 1 層の全結合層のみからなる Network を学習した際の結果と, 学習前の初期化された状態の PyramidNet を用いて全結合層のみを学習した結果を併記している.

表 2 の結果から, 関連性が高くなるに従って分類精度が向上することを確認した. 驚くべきことにランダムなラベルを学習した場合について, 学習前や全結合層のみの場合に比べ認識精度が向上することを確認した. この結果は本来の分類対象に比べて関連が低い場合も学習に寄与することを示唆している. またラベルに関わらず何らかの分類問題を学習すること自体が分類精度に寄与する可能性を示している.

更にこのモデルが画像分類の学習に寄与するか, 転移学習を用いて評価した. 結果を表 3 に示す. 表 3 の結果から, 転移学習においては表 2 の差が埋まり, 認識精度に大きな差が見られなくなることが分かった. また, 僅かながらスーパークラスを用いた際の分類精度がランダム初期値を上回った. このことから, スーパークラスを追加タスクとして用いた際に, 認識精度の改善が期待できると考えられる.

4.2 実験 2

Data Augmentation を推定する追加タスクについて同時学習を行い, 追加タスクによる精度向上が起きるかを評価した. 特徴抽出部として ResNet [16] の全結合層直前までを用いた構

表 4: Data Augmentation に対する追加タスクのみで学習した際の最終 Epoch の Accuracy.

Data Augmentation	Train Accuracy
Random Crop	82.35%
cutout	99.95%
RICAP	69.85%

表 5: Data Augmentation に対する追加タスクと本来の分類タスクを同時に扱った際の最終 Epoch の Accuracy. ただしそれぞれの結果は 4 回の平均および標準偏差を示している.

Data Augmentation	追加タスク	Test Accuracy
Random Crop	-	73.39% \pm 0.39%
	○	73.57% \pm 0.35%
cutout	-	76.02% \pm 0.50%
	○	76.18% \pm 0.30%
RICAP	-	76.30% \pm 0.59%
	○	76.52% \pm 0.45%

造を利用した. ResNet は層数 110 で Basic Block を用いたものを用いた. cutout は Random Crop と Horizontal Flip の後に中央の位置を一様乱数に基づき決定し, 16×16 の領域を対象として切り取った. RICAP は 4 枚の画像に Random Crop と Horizontal Flip を施した後に 4 枚の画像の接合点をベータ分布 $Beta(0.3, 0.3)$ の乱数に基づき決定し, 画像を貼り合わせた.

それぞれの追加タスクに対してのみ学習を行った際の Training Accuracy を表 4 に示す. cutout に対してのみ適合することが分かった. ただし, Random Crop は Horizontal Flip を併用しない場合は適合することを確認した. また, RICAP は 4 枚の画像の接合点が画像の左端/右端, 上端/下端に存在する可能性があり, このとき上下や左右に並んだ 2 枚, 或いは双方を満たした場合には 1 枚の画像として現れる. このとき, Segmentation で 4 パターンのうちのどの画像が出現したか区別することが出来ないため, 失敗してしまう. 接合点が端に現れないように設定した場合は適合することを確認した.

追加タスクと本来の分類タスクを同時に扱った際の実験結果を表 5 に示す. いずれの追加タスクにおいても分類精度は改善したが, 大きな改善には至らなかった.

4.3 考察

実験 1, 実験 2 の結果の双方で, 分類精度の改善は達成できなかった. 原因として Epoch 数の多さと, 今回対象とした追加タスクの容易さが挙げられる.

He らの研究では, 学習回数に応じて ImageNet の事前学習を行った場合とランダムな値で初期化を行った際の精度の差が縮むことが示されている [17]. 従って, 学習回数が十分大きい場合には, 転移学習元の特徴表現能力に関わらず同様の分類精度になる可能性がある. 学習初期の Training Loss および Training Accuracy の推移を見ることで, これは検証できる. 表 3 の転移学習実験における, 冒頭 50 Epoch の Training Loss を図 2, Training Accuracy を図 3 に示す. 図中には比較のため, ランダム初期値から学習を行った際の結果を併記している.

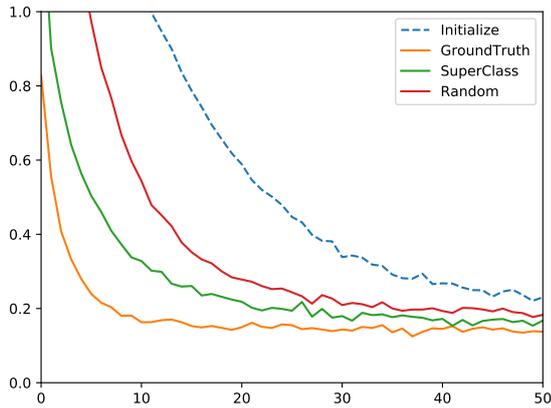


図 2: 転移学習実験における学習冒頭 50 Epoch の Training Loss の推移. 比較のため, 初期値からの学習の推移を点線で示している. Ground Truth, スーパークラス, ランダム, 初期値の順に学習データに対して適合しており, 学習が進むにつれ各条件の Loss が漸近していることが観測できる. これは表 2 で示された分類精度順になっている.

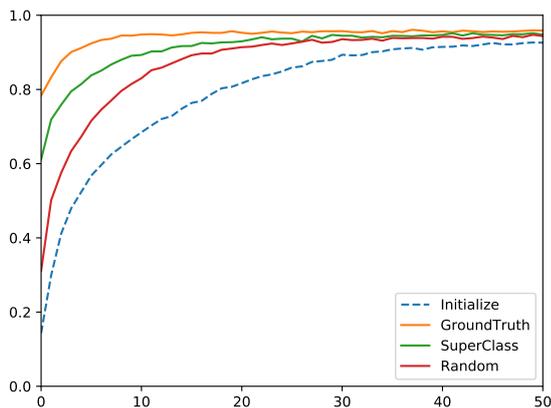


図 3: 転移学習実験における学習冒頭 50 Epoch の Training Accuracy の推移. 比較のため, 初期値からの学習の推移を点線で示している. Loss と同様に Accuracy でも Ground Truth, スーパークラス, ランダム, 初期値の順に学習データに対して適合しており, 学習が進むにつれ各条件の Accuracy が漸近していることが観測できる. これは表 2 で示された分類精度順になっている.

図 2, 図 3 はそれぞれの学習条件が冒頭 50 Epoch の段階ではほぼ同じになることを示しており, 今回の実験に関しては Epoch 数の多さから, 分類精度が一様になってしまった可能性がある.

また, 今回対象とした追加タスクは, いずれも条件によって学習データに対して容易に適合してしまうタスクであった. これは追加タスクの学習が容易であり, 画像分類タスクに悪影響を及ぼさない反面, 好影響をもたらさないほどに容易であった可能性がある. より大幅な改善を目指すのであれば, より高難易度の追加タスクを検討する必要があると考えられる.

5. まとめと今後の課題

本来の画像分類とは異なる Pretext Task の成功を背景に, 異なる関連性のラベルの画像分類を Pretext Task の代替として用いることで, どのような視覚的能力が目的の画像分類タスクに寄与するのか分析を行った. 実験結果からは, 全くランダムに与えられたラベルの学習であっても, 本来の画像分類タスクに寄与することを確認した. ただし, 我々が検討した追加タスクは認識精度を改善しなかった.

今後は特徴の可視化や重みの詳細な分析を通して, 学習タスクと目的の画像分類との関連性を詳細に分析する. また, 異なるモデルやデータセットを用いて同様の分析を行う.

謝辞 本研究は, JST CREST #JPMJCR16E1, JSPS 科研費#17H01803 と#18K19785 の補助による.

文 献

- [1] A. Krizhevsky, I. Sutskever, and G.E. Hinton, “ImageNet classification with deep convolutional neural networks,” Proc. NIPS, 2012.
- [2] Y. Huang, Y. Cheng, D. Chen, H. Lee, J. Ngiam, Q.V. Le, and Z. Chen, “Gpipe: Efficient training of giant neural networks using pipeline parallelism,” arXiv preprint arXiv:1811.06965, 2018.
- [3] T. Miyato, S.-I. Maeda, S. Ishii, and M. Koyama, “Virtual adversarial training: A regularization method for supervised and semi-supervised learning,” IEEE Trans. PAMI, 2018.
- [4] M. Caron, P. Bojanowski, A. Joulin, and M. Douze, “Deep clustering for unsupervised learning of visual features,” Proc. ECCV, 2018.
- [5] S. Gidaris, P. Singh, and N. Komodakis, “Unsupervised representation learning by predicting image rotations,” Proc. ICLR, 2018.
- [6] M. Noroozi, A. Vinjimoor, P. Favaro, and H. Pirsiavash, “Boosting self-supervised learning via knowledge transfer,” Proc. ACML, 2018.
- [7] M. Noroozi, H. Pirsiavash, and P. Favaro, “Representation learning by learning to count,” Proc. ICCV, 2017.
- [8] M. Noroozi and P. Favaro, “Unsupervised learning of visual representations by solving jigsaw puzzles,” Proc. ECCV, 2016.
- [9] J. Frankle and M. Carbin, “The lottery ticket hypothesis: Finding sparse, trainable neural networks,” arXiv preprint arXiv:1803.03635, 2018.
- [10] C. Zhang, S. Bengio, M. Hardt, B. Recht, and O. Vinyals, “Understanding deep learning requires rethinking generalization,” Proc. ICLR, 2017.
- [11] R. Geirhos, P. Rubisch, C. Michaelis, M. Bethge, F.A. Wichmann, and W. Brendel, “Imagenet-trained cnns are biased towards texture; increasing shape bias improves accuracy and robustness,” arXiv preprint arXiv:1811.12231, 2018.
- [12] L.E.A.T.A.M. Dimitris Tsipras, Shibani Santurkar, “Robustness may be at odds with accuracy,” arXiv preprint arXiv:1805.12152, 2018.
- [13] T. DeVries and G.W. Taylor, “Improved regularization of convolutional neural networks with cutout,” arXiv preprint arXiv:1708.04552, 2017.
- [14] R. Takahashi, T. Matsubara, and K. Uehara, “Ricap: Random image cropping and patching data augmentation for deep cnns,” Proc. ACML, 2018.
- [15] D. Han, J. Kim, and J. Kim, “Deep pyramidal residual networks,” Proc. CVPR, 2017.
- [16] K. He, X. Zhang, S. Ren, and J. Sun, “Deep residual learning for image recognition,” Proc. CVPR, 2016.

- [17] K. He, R. Girshick, and P. Dollr, “Rethinking imagenet pre-training,” arXiv preprint arXiv:1811.08883, 2018.