

深層学習における敵対的ネットワークを用いた 漫画画像の自動カラー化

片岡 裕介^{†1} 松原 崇^{†1} 上原 邦昭^{†1}

概要：近年スマートフォンやタブレット端末と行った電子機器が発達し、雑誌で掲載されていた漫画を電子書籍で読むことが可能となっている。また雑誌を印刷する必要がなくなるために、カラー化した画像を使用しても印刷のコストが掛からなくなった。そこで従来の白黒で描かれた漫画画像をデジタル上で編集し、カラー化したものを電子書籍で扱うことが考えられる。この時、自動カラー化技術を用いれば、漫画をカラーにすることができれば、人的コストや作業時間を大幅に減らすことが可能となる。本研究では、深層学習を用いた漫画画像のカラー化について検討する。本システムは、モノクロの入力画像から、輪郭などの特徴情報を抽出するネットワークと、カラーの入力画像から色情報を抽出するネットワークからなる。この2つのネットワークの出力を、色付けを行うネットワークの入力とし、カラー画像を出力する。本システムは、敵対的ネットワークを用いた学習によって、自然なカラー化を可能とする点に特徴がある。また、色情報を抽出するネットワークの出力の分布が、事前分布になるように学習を行うことにより、事前分布からサンプリングした値に従った任意の色付けが可能であり、より自然なカラー化が可能となっている。

KATAOKA YUUSUKE^{†1} MATSUBARA TAKASHI^{†1} UEHARA KUNIAKI^{†1}

1. はじめに

近年スマートフォンやタブレット端末と行った電子機器が発達し、雑誌で掲載されていた漫画を電子書籍で読むことが可能となっている。この際、印刷によるインクの消費がなくなるために、カラー画像を使用することによるコストを考慮する必要はなくなっている。そこで従来の白黒で描かれた漫画画像をデジタル上で編集し、カラー化したものを電子書籍で扱うことが考えられる。この時、自動カラー化技術を用いれば、漫画をカラーにすることができ、人的コストや作業時間を大幅に減らすことが可能となる。

白黒画像をカラー画像にするという研究はコンピュータービジョンの分野において盛んに行われている。特に、自然画像のカラー化においては、白黒写真や白黒映画をカラー化するというモチベーションで、様々な手法が試されている。その手法の1つとして、深層ニューラルネットワークを用いて、画像のカラー化をおこなうものがある [1]。これは、モノクロ画像からカラー画像を復元するように、大量のデータを用いてネットワークに学習させて、画像

の自動カラー化を行うというもので、白黒画像に対して、風景の画像の自然なカラー化が可能としている。また、学習済み深層ニューラルネットワークを用いた手法では [2]、ネットワークの隠れ層の出力から、色の彩度と明るさを復元するよう学習を行い、様々な対象が写った画像についても自然なカラー化が可能となっている。また、深層畳み込みニューラルネットワークを用いた手法 [3] では、グレースケール画像をネットワークの入力とし、大域特徴量と局所特徴量を抽出して、全体の色合いと部分的な色合いを考慮した色付けが可能となっている。また、大域特徴量を用いたクラス分類が可能となっている。しかし、これらの研究では、カラー化された画像中の物体の色が、目的の色にならないという問題点があった。これは、ネットワークが色情報を扱う潜在変数を推論をする構造ではないからである。また従来の研究のカラー化の対象の多くは自然画像である。自然画像では、グレースケール画像の濃淡情報に対し、色彩情報を付加すればカラー画像を復元させることが可能である。それに対し、漫画画像はスクリーン Tone や特殊な描画方法を用いて濃淡を表現している他、肌の色や髪の毛の色などが白一色に統一されているなど、自然画像のように、グレースケールのデータに色情報を付加する

^{†1} 現在、神戸大学大学院システム情報学研究所
Presently with Graduate School of System Informatics, Kobe University

だけではカラー化することはできない。また、深層ニューラルネットワークの学習には、大量のデータセットが必要になるが、漫画画像は対応するカラー画像のデータが少なく、カラー化の学習に使用できるデータの数が少ないという問題点がある。

そこで、画像にエッジ検出法等の前処理を行い、イラストやアニメ画像、漫画画像から同じような画像を生成できる特徴情報を抽出し、グレースケール情報を含めた色情情報復元の学習を行う。こうすることで、大量のイラストやアニメ画像を学習データに使用しても、漫画画像のカラー化を学習させることが可能になると考えられる。また、色彩画像から潜在変数を推論し、潜在変数が事前分布に従うように学習を行う。こうすることで潜在変数は色情情報を扱えるようになり、潜在変数の値を変えることで必要な色で画像の着色を行うことができる。さらに、与えられた潜在変数に従った着色が可能になると考えられる。

データの生成と潜在変数の推論を行う手法としては、変分オートエンコーダを用いた手法 [4] が挙げられる。変分オートエンコーダは、変分ベイズ法に基づき画像と潜在変数の推論の学習を行う。これにより、潜在変数の値を変えれば、生成画像を変更することが可能になる。しかし変分オートエンコーダは画像データについて画素単位の誤差に対して学習を行うため、輪郭がぼやけた画像などの不自然な画像が生成される問題点がある。それに対し、敵対的ネットワーク [5] を用いた手法では、ヒューリスティックな誤差に基づいた学習を行うため、画素単位の誤差による学習よりも自然な画像を生成することが可能である。

特に深層畳み込み敵対的ネットワーク [6] では、多層の畳み込みニューラルネットワークを敵対的ネットワークで学習することで、人間の顔画像等の生成で高い成果を挙げている。また、条件付き敵対的ネットワーク [7], [8] では、条件データを考慮して、与えられた条件データに従った画像を生成することが可能である。特に、条件付き敵対的ネットワークを用いたカラー化の手法 [9] では、エッジ画像やセマンティックセグメンテーション処理済みの画像などを入力にしてカラー画像を生成することを可能としている。

しかしこれらの手法では、データから潜在変数を推論することが出来ないという問題点がある。このため、データと潜在変数の学習法として敵対的ネットワークによる学習と、変分オートエンコーダを用いた学習を同時に行う手法が提案されている [10]。この手法では、通常の敵対的ネットワークや変分オートエンコーダよりも、自然な画像を生成でき、また、潜在変数の推論も可能である。しかし、変分オートエンコーダを用いるために、画像がぼやけるという問題点がある。敵対的ネットワークを用いたデータ空間と潜在変数の推論の同時学習法 [11], [12] では、ヒューリスティックな誤差のみによる学習が可能であり、また潜在変数の推論も可能である。

本研究では、イラスト画像にエッジ検出のような前処理を行い、特徴画像を学習データとして用いた。これにより、漫画のような濃淡表現を行っている画像に対しても、特徴画像からカラー化が可能であることを示す。また、特徴画像を条件データとした敵対的ネットワークを用いたカラー化画像と色情情報の潜在変数の推論の同時学習法を用い、特徴画像を考慮することにより、自然なカラー化ができて、また外部から与えた色情情報に従ったカラー化が可能であることを示す。

2. 関連研究

2.1 条件付き敵対的ネットワーク

敵対的ネットワークを拡張した学習法として、条件 y を考慮した敵対的ネットワークによる学習法が挙げられる [7], [8]。通常の敵対的ネットワークは、ランダムな潜在変数 z からデータ x を出力する生成部 G と、元データ x と生成部が出力したデータ $x = G(z)$ を入力にとり、各入力が元データである確率を出力する識別部 D から構成される。

条件付き敵対的ネットワークでは G と D の入力に、条件データ y を加える。 G は $D(G(z, y))$ を最大化するように学習を行い、同時に D は $D(x, y)$ と $1 - D(G(z, y), y)$ を最大化するように学習を行う。この時、 G と D はそれぞれ式 (1) と式 (2) を最大化するように学習を行う。

$$L_G = \log D(G(z, y)) \quad (1)$$

$$L_D = \log D(x, y) + \log(1 - D(G(z, y), y)) \quad (2)$$

y を G と D の入力に加えるために、 G は y の値に従ったデータの生成を行うように学習が進み、 D は y の値に従って、元データ x が入力された時に出力する確率が大きくなるように学習が進む。学習が収束すると、 D は x とそれ以外のデータを識別するようになり、 G は D が x と識別するような、自然なデータを生成するようになる。

2.2 敵対的ネットワークを用いたデータ空間と潜在変数の同時学習

敵対的ネットワークを用いた、データ空間と潜在変数の推論の同時学習法 [11] では、元データ x から潜在変数を出力する推論部 G_z と、潜在変数 z からデータを出力する生成部 G_x 、識別部 D から構成される。 D は $(x, G_z(x))$ と $(G_x(z), z)$ を入力に取り、各入力が推論部 G_z の入力と出力の同時分布である確率を出力する。この時、 G_z と G_x は、 $1 - D(x, G_z(x))$ と $D(G_x(z), z)$ を最大化するように学習を行い、 D は $D(x, G_z(x))$ と $1 - D(G_x(z), z)$ を最大化するように学習を行う。この時、 G_z と G_x 、 D はそれぞれ式 (3) と式 (4) を最大化するように学習を行う。

$$L_G = \log(1 - D(x, G_z(x))) + \log D(G_x(z), z) \quad (3)$$

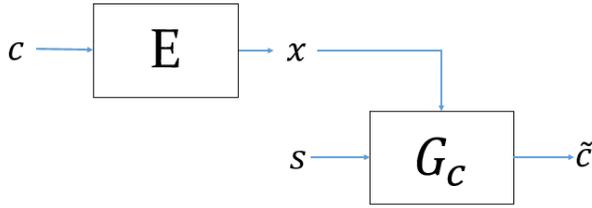


図 1 色付けネットワーク G_c のモデル図

$$L_D = \log D(x, G_z(x)) + \log(1 - D(G_x(z), z)) \quad (4)$$

G_z は入力された x に従った潜在変数を推論するように学習が進み, G_x は入力された z からより本物に近い自然なデータを生成するように学習が進む. D は $(x, G_z(x))$ が入力に与えられた時の出力の確率値が大きくなるように学習を進める. 学習が収束すると, $G_x(z)$ はデータ x に似ている自然な画像を生成するようになり, $G_z(x)$ は z が取りうる値の範囲に収まるようになる.

3. 提案手法

3.1 条件付き敵対的ネットワークを用いたデータ空間と潜在変数の同時学習

本手法では, 条件データ付き敵対的ネットワークを用いた, データ空間と潜在変数の推論の同時学習法を提案する. 条件付き敵対的ネットワークとは, 通常の敵対的ネットワークの拡張であり, 条件データを用いた学習を行う. この条件データを, 敵対的ネットワークを用いた, データ空間と潜在変数の同時学習を行う際にも使用するのが, 提案手法の特徴である. 本手法では, ネットワークは色情報抽出ネットワーク G_s と色付けネットワーク G_c , 識別部 D から構成される. G_c は特徴画像と潜在変数からカラー化画像の生成を行い, G_s はカラー画像から色情報の潜在変数を出力する. G_c のネットワークモデルを図 1 に示す. 学習に使用するカラー画像 c から抽出した特徴画像を x とする. ここで, 色の特徴量 s と x を G_c の入力として, カラー化画像 \tilde{c} を生成する.

$$x = E(c) \quad (5)$$

$$\tilde{c} = G_c(x, s) \quad (6)$$

また, c を G_s の入力として, 色の特徴量 $s_{inferred}$ を出力する. ここで, $s_{inferred}$ を元に画像のカラー化を行う場合は, G_c の入力に $s_{inferred}$ を使用する. G_s のネットワークモデルを図 2 に示す.

$$s_{inferred} = G_s(c) \quad (7)$$

$$\tilde{c}_{recolored} = G_c(x, s_{inferred}) \quad (8)$$

D のネットワークモデルを図 3 に示す. D は x と c , s を入力に取り, 確率 d を出力する. d は入力された c と s が,

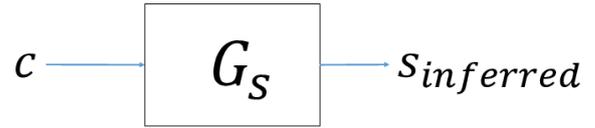


図 2 色情報抽出ネットワーク G_s のモデル図



図 3 識別部 D のモデル図

条件として y を持つ時に, G_c の入力と出力のペアである確率を表している.

3.2 敵対的ネットワークによる評価関数の定義

敵対的ネットワークを用いた学習をするために, x と事前分布 p からサンプリングした特徴量 $s_{sampled}$ を G_c の入力にして, カラー化画像 $\tilde{c}_{painted}$ を生成する.

$$s_{sampled} \sim p(s) \quad (9)$$

$$\tilde{c}_{painted} = G_c(x, s_{sampled}) \quad (10)$$

D は $(x, s_{inferred}, \tilde{c}_{recolored})$ と $(x, s_{sampled}, \tilde{c}_{painted})$ を入力に取り, それぞれについて d を出力する. この時, G_c と G_s は式 (11) を最大化するように学習を行い, D は式 (12) を最大化するように学習を行う.

$$L_G = \log(1 - D(x, c, s_{inferred})) + \log D(x, \tilde{c}_{painted}, s_{sampled}) \quad (11)$$

$$L_D = \log(D(x, c, s_{inferred})) + \log(1 - D(x, \tilde{c}_{painted}, s_{sampled})) \quad (12)$$

G_c は x の値に従って, s の色情報に基づいたカラー化画像を生成するように学習が進み, G_s は c から, 正しい色情報 s を推論するように学習が進む. D は $(x, c, s_{inferred})$ と $(x, \tilde{c}_{painted}, s_{sampled})$ を識別するように学習を行う. 学習が収束すると c から推論された $s_{inferred}$ は事前分布 p に従うような値を取るようになる. また, $s_{sampled}$ から生成された $\tilde{c}_{painted}$ は c と区別がつかないような, 自然な画像になる.

3.3 画素単位の誤差による学習

敵対的ネットワークを用いた学習に加え, 画素単位の誤差として平均二乗誤差 $RMSE$ を用いた学習も行う. 画素

単位の誤差を導入することで、画像中の対象の適切な箇所、カラー化が行われやすくなる。この時 G_c と G_s は式 (13) を最小化するように学習を行う。

$$L_{pixelwise} = RMSE(c, \tilde{c}_{recolored}) \quad (13)$$

この時、 G_c は生成した $\tilde{c}_{recolored}$ が c に近づくように学習を行うため、より元画像の色情報に近いカラー化画像を生成するようになる。また、 G_s は c から推論した $s_{inferred}$ がより自然に着色できるような潜在変数になるように学習が進む。

3.4 敵対的ネットワークと画素単位の誤差を組み合わせた学習

式 (11) と式 (13) から、 G_c と G_s は式 (14) を最適化するように学習を行う。 D の評価関数は式 (12) と同様である。

$$\min L = L_{pixelwise} - LG \quad (14)$$

敵対的ネットワークを用いた最適化により、カラー化画像と色情報の潜在変数と推論を同時に行うことができ、また、画素単位の誤差を用いた最適化により、画像の適切な箇所にカラー化されるように学習が進むと考えられる。

4. 評価実験

4.1 実験設定

評価実験にはテスト用に Nico-illustrate データセット*1を利用した。学習用に Web 上からクロールして集めたカラーのイラスト画像 300 万枚を使用した。それぞれ各画像に縮小、切り取りの前処理を行い入力サイズを固定する。本実験では、画像サイズ 256 × 256 にデータを整形し、学習とテストを行った。関数 E によって画像の特徴情報を抽出する手法としてはエッジ検出の手法である canny 法を使用した。またテストでは、カラー画像と漫画調の白黒画像について、以下の条件で実験を行った。ただし白黒画像については条件 (1) で実験は行わない。

- (1) 元となるカラー画像 c の色情報を用いたカラー化
 - (2) ランダムな色情報 $s_{sampled}$ によるカラー化
 - (3) 異なるカラー画像 c' の色情報を用いたカラー化
- 条件 (1) の結果から、元のカラー画像が存在する画像についても、その色情報に従った着色が可能であるか検証できる。条件 (2) の結果から、ランダムにサンプリングした色情報を用いても自然な着色が可能であるか検証できる。条件 (3) の結果から、異なるカラー画像の色情報を用いた場合でも、その色情報に従った着色が可能であるか検証できる。この時、 c' を G_c の入力にすることで、カラー画像 $\tilde{c}_{transferred}$ を生成する。

$$s_{extravted} = G_s(\tilde{c}') \quad (15)$$

*1 <https://nico-opendata.jp/>

$$\tilde{c}_{transferred} = G_c(x, s_{extravted}) \quad (16)$$

$s_{extravted}$ は c' より推論された色情報の潜在変数である。適切に学習が進んでいけば、 $\tilde{c}_{transferred}$ は c' の色情報を考慮したカラー化画像になると考えられる。

4.2 カラー画像の再カラー化

カラー画像から抽出した特徴画像から画像を再カラー化した結果を図 4 に示す。左側の列から、それぞれ元のカラー画像 c と特徴画像 x 、条件 (1) に従って生成した \tilde{c} 、条件 (2) に従って生成した $\tilde{c}_{painted}$ 、条件 (3) に従って生成した $\tilde{c}_{transferred}$ を表している。 $\tilde{c}_{recolored}$ と c を比較すると、元のカラー画像の色情報を考慮した着色が出来ていることが分かる。また、 $\tilde{c}_{painted}$ と c を比較すると、ランダムにサンプリングした色情報を用いても、自然な着色が出来ていることが分かる。また、 $\tilde{c}_{transferred}$ と c 、 c' を比較すると、異なるカラー画像の色情報を用いても、その色情報を考慮した着色が可能であることが分かる。また、 $\tilde{c}_{recolored}$ と $\tilde{c}_{painted}$ 、 $\tilde{c}_{transferred}$ を比較すると、各部位を考慮した色付けが行っていることが分かる。このことから色情報の潜在変数は、キャラクターの各部位に対応する色の情報を持っていると考えられる。

4.3 白黒画像のカラー化

カラー画像から抽出した特徴画像から画像を再カラー化した結果を図 5 に示す。左側の列から、それぞれ元の白黒画像 c と特徴画像 x 、条件 (2) に従って生成した $\tilde{c}_{painted}$ 、条件 (3) に従って生成した $\tilde{c}_{transferred}$ を表している。 $\tilde{c}_{painted}$ を見ると、ランダムにサンプリングした色情報を用いても、自然な着色が出来ていることが分かる。また $\tilde{c}_{transferred}$ と c' を比較することで、異なるカラー画像の色情報を用いても、その色情報を考慮した着色が可能であることが分かる。

5. 結論

本研究では、特徴情報を条件とした敵対的ネットワークによる、カラー画像の生成と色情報の推論の同時学習法を用いた白黒の漫画画像のカラー化手法を提案した。評価実験には Nico-illustrate データセットを使用し、カラー画像と漫画調の白黒画像について画像のカラー化ができることを示した。カラー画像において、カラー画像の色情報を与えることで、元のカラー画像のように着色ができることを示した。また、ランダムに生成した色情報を与えても、自然なカラー化が可能であることを示した。また、異なるカラー画像から抽出した色情報を与えても、カラー化が可能であることを示した。

今後の課題として、カラー画像と白黒画像からより自然かつ同一の形式となる特徴情報を抽出する手法を適用することが挙げられる。canny 法を用いた特徴抽出では、カ

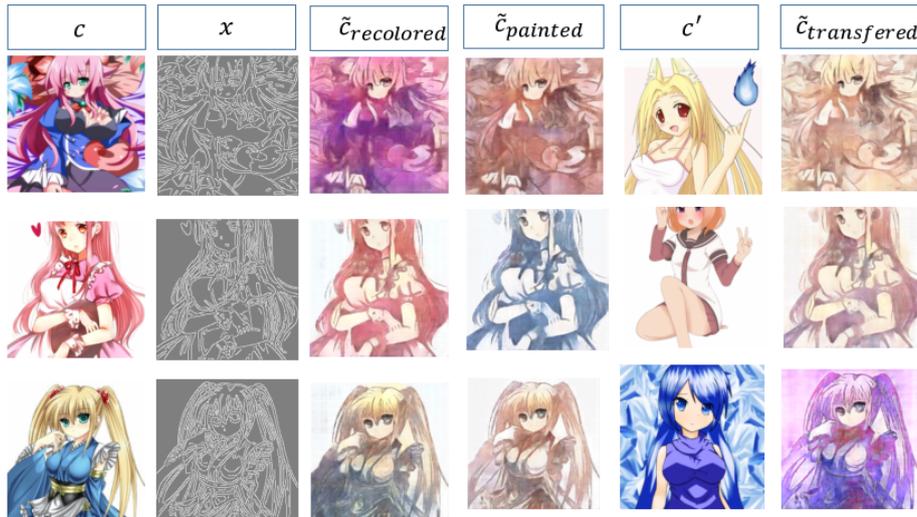


図 4 色情報をからカラー画像を再カラー化

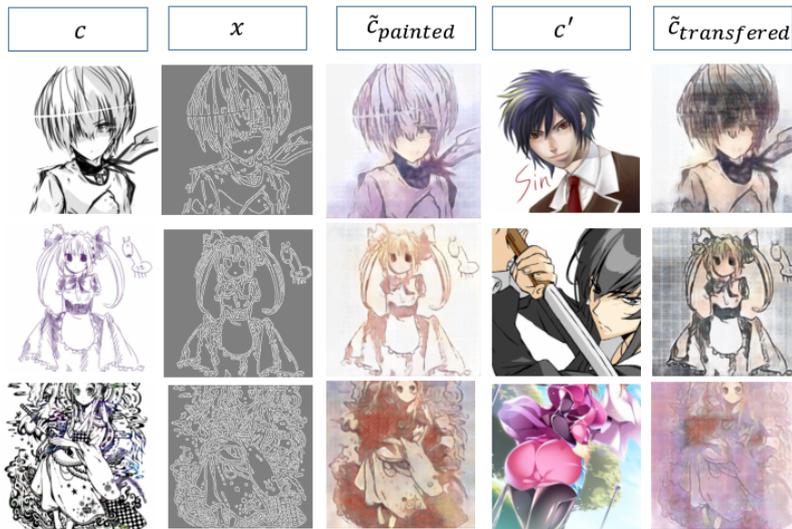


図 5 色情報をから白黒画像を再カラー化

ラー画像に適用した場合と白黒画像に適用した場合とで、出力結果が少し異なる。その結果として白黒画像から抽出した特徴情報では学習に用いた特徴情報とデータの種類が変わることで、画像の色付けが不十分になると考えられる。また、今回はカラー化画像した画像の定性的な評価のみだったが、定量的な評価手法による実験結果の比較などが挙げられる。

謝辞

本研究は科学研究費補助金 26280040 および 16K12487 の補助を受けて実施された

参考文献

[1] Cheng, Z., Yang, Q. and Sheng, B.: Deep colorization, *Proceedings of the IEEE International Conference on*

Computer Vision, pp. 415–423 (2015).
 [2] Larsson, G., Maire, M. and Shakhnarovich, G.: Learning representations for automatic colorization, *European Conference on Computer Vision*, Springer, pp. 577–593 (2016).
 [3] Iizuka, S., Edgar Simo-Serra and Ishikawa, H.: Let there be Color!: Joint End-to-end Learning of Global and Local Image Priors for Automatic Image Colorization with Simultaneous Classification, *Siggraph '16* (2016).
 [4] Kingma, D. P. and Welling, M.: Auto-encoding variational bayes, *arXiv preprint arXiv:1312.6114* (2013).
 [5] Goodfellow, I., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., Courville, A. and Bengio, Y.: Generative adversarial nets, *Advances in neural information processing systems*, pp. 2672–2680 (2014).
 [6] Radford, A., Metz, L. and Chintala, S.: Unsupervised representation learning with deep convolu-

- tional generative adversarial networks, *arXiv preprint arXiv:1511.06434* (2015).
- [7] Gauthier, J.: Conditional generative adversarial nets for convolutional face generation, *Class Project for Stanford CS231N: Convolutional Neural Networks for Visual Recognition, Winter semester 2014* (2014).
 - [8] Mirza, M. and Osindero, S.: Conditional generative adversarial nets, *arXiv preprint arXiv:1411.1784* (2014).
 - [9] Isola, P., Zhu, J.-Y., Zhou, T. and Efros, A. A.: Image-to-image translation with conditional adversarial networks, *arXiv preprint arXiv:1611.07004* (2016).
 - [10] Larsen, A. B. L., Sønderby, S. K., Larochelle, H. and Winther, O.: Autoencoding beyond pixels using a learned similarity metric, *Proceedings of The 33rd International Conference on Machine Learning*, pp. 1558–1566 (2016).
 - [11] Dumoulin, V., Belghazi, I., Poole, B., Lamb, A., Arjovsky, M., Mastropietro, O. and Courville, A.: Adversarially learned inference, *arXiv preprint arXiv:1606.00704* (2016).
 - [12] Donahue, J., Krähenbühl, P. and Darrell, T.: Adversarial feature learning, *arXiv preprint arXiv:1605.09782* (2016).