Vol.2022-CVIM-229 No.34
2022/3/11

教師なし画像生成における色情報制約を用いたドメイン適応

小林　賢也[†1]　和田　直哉[†1]

概要：近年, 画像変換のタスクにおいて様々な研究が精力的に行われており, 特に Generative Adversarial Network（GAN）が注目されている．本研究では，人体のシミュレータ画像と実画像の画像変換において，特定の人物の皮膚色・明るさに合わせたドメイン適応手法を提案する．具体的には，少数かつペアでないデータで学習可能な CycleGAN に色分布制約を導入したネットワークモデルを用いて，MRI で得られた膝の 3D モデル画像群・膝の外観画像群の双方を学習する．色分布制約にはカーネル密度推定を用いて算出した生成画像の色分布ヒストグラムとリファレンス画像の色分布ヒストグラムとの L2 損失を適用する．提案手法は OAI（OsteoArthritis Initiative）の膝のデータセットを用いた検証実験によって，少数のペアでないデータを用いて，特定の人物の皮膚色・明るさを再現する多様な画像生成を可能にしたことを確認した．

キーワード：GAN，スタイル変換，ドメイン変換，色分布ヒストグラム

1. はじめに

本論文では，ドメイン変換タスクにおいて，ドメイン内の色分布に着目し，特定の色を再現するドメイン変換を実現することを目的とする．

近年, 画像生成技術が注目されており, 中でも Generative Adversarial Network（GAN）[1，9，10，11]は高品質な画像を生成できることが知られている．また，GAN は異なるドメイン間の画像変換のタスクにおいても素晴らしい結果が出ることが分かっている[2，3]．一方で，ドメイン内には色や照明条件など，複数の「小さい」ドメインが存在し，実用段階においてはそれを再現したい場合がある．しかしながら，ドメイン内の色分布から特定の色を反映する画像生成タスクは困難であった．その原因として，画像の色分布を表すヒストグラムの計算は微分可能ではなく，深層学習のようなフレームワークに導入できない問題があり，生成画像の色分布を制御できないことが挙げられる．

本研究では，この問題を解決するために, CycleGAN[2]のネットワークモデルに Deephist[4]で提案されたカーネル密度推定を用いた色情報制約を導入したネットワークモデルを提案する．これにより，少数かつペアでないデータを用いた生成画像の色分布制御が可能になると考える．

検証実験では，アメリカの国立衛生研究所（National Institutes of Health）から支援を受ける OAI（OsteoArthritis Initiative）が提供する膝の MRI データセットと Kaggle が提供する Fashion Product Images Dataset から抽出した膝の実画像のデータを用いて，提案したネットワークモデルがドメイン間のスタイル変換において，ドメイン内の色分布の制御を可能にすることを示す．

2. 関連研究

GAN [1，9，10，11]はデータ分布を捉えて，あるドメインへのスタイル変換タスクによく利用されてきた[2，3，6，7，8]．Pix2Pix[3]は，その代表的な手法の一つで，対になったデータを用いて教師あり学習を行い，より詳細で鮮明なスタイル変換を可能にした．しかしながら，対になるデータを大量に取得するのは困難であり，課題であった．その課題を解決したのが CycleGAN である. CycleGAN は cycle-consistency loss という一貫性損失を導入することによって，少数かつ不対データを用いた教師なし学習を可能にした．このようなスタイル変換モデルはあるドメインへの変換において，データ分布のモードに依存してしまい，明示的に生成画像の色分布を制御することは難しい．

最近では，ドメイン内の多様な特徴を表現するスタイルコード自体を学習する StarGAN-v2[5]が提案されている. このネットワークアーキテクチャは同一ドメイン内における多様な特徴を生成画像に反映することを可能にした. また，ドメイン内の特徴である色分布を制御する手法として，Deephist が提案されている．Deephist はカーネル密度推定を用いて生成画像から色分布ヒストグラムを推定することによって，微分可能なネットワーク構造を構築した．そして，得られた生成画像の色分布ヒストグラムとリファレンス画像の色分布ヒストグラムとの損失関数として EMD(Earth Mover’s Distance)を導入し, 生成画像の色分布の制御を可能にした．

本研究と関連研究の相違点に関して, StarGAN-v2 はドメイン内の多様な特徴を表現するスタイルコードの学習のために，大量の学習データが必要になる．また，Deephist はペアな画像を複数必要とする半教師学習のような形を取る．一方，我々の手法は少数かつペアでないデータを用いて，

†1 京セラ(株)
Kyocera Corp.

ドメイン内の多様な色分布を生成画像に反映させるネットワークモデルを提案する．

3. 提案手法

我々は CycleGAN に Deephist の色情報制約を導入したネットワークモデルを構築する．CycleGAN の持つ cycle-consistency 損失などの特性により，少数かつペアでないデータ学習で形状不変なスタイル変換を実現し，Deephist の色情報制約損失により，同一ドメイン内の色分布の自在な制御を可能にすると考える．この章ではネットワークモデルの構造について詳細に述べる．

3.1 ネットワーク構造

図 1 に示すようにネットワークモデル全体は二つの生成器と識別器から構成される．具体的には，ドメイン X から Y への変換を行う生成器G，ドメイン X の真贋判定を行う識別器D_{X}，ドメイン Y から X への変換を行う生成器F，ドメイン Y の真贋判定を行う識別器D_{Y}から構成される．生成器G，Fは中間層にリファレンス画像X_{ref}，Y_{ref}の色分布ヒストグラムX_{hist}，Y_{hist}を入力するような構造を取る．ここで，生成器，識別器には GAN の安定化のため，Spectral Normalization[12]を導入した．3.3 節で生成器の詳細な構造を述べる．

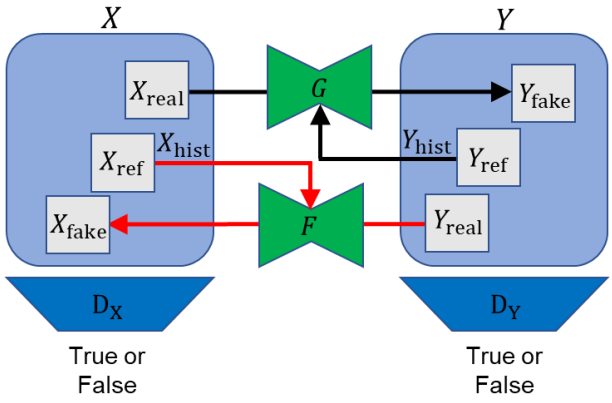

図 1 ネットワーク概略図

3.2 色分布制約

リファレンス画像から RGB の各色分布ヒストグラムを算出し，その色分布ヒストグラムから抽出した色を表現する特徴量を生成器の中間層に与える．また，生成画像の色分布ヒストグラムとリファレンス画像の色分布ヒストグラムを比較する色情報における損失関数を導入する．

まず，生成画像から微分可能な色分布ヒストグラムを推定するために，Deephist が用いたカーネル密度推定（Kernel Density Estimation: KDE）を利用する．以下の式は色分布ヒストグラムを推定するための確率密度関数である．

$$f_I(g) = \frac{1}{NB}\sum_{x\in\Omega} \mathrm{K}\left(\frac{I(x)-g}{B}\right) \tag{1}$$

ここで，xは画像ピクセル位置，$g \in [-1,1]$は画像の画素値レベル，$\mathrm{K}(\cdot)$はカーネル関数，Bはバンド幅，$N = |\Omega|$は画像のピクセル数，$I(x) \in [-1,1]$は画像ピクセル位置xの輝度値を表す．式(1)のカーネル関数は以下のように定義する．

$$\mathrm{K}(z) = \frac{d}{dz}\sigma(z) = \sigma(z)\sigma(-z) \tag{2}$$

カーネル関数$\mathrm{K}(z)$はシグモイド関数$\sigma(z)$の導関数を表す．正規化された画素値$[-1,1]$をK区間に分割し，各々の区間の長さ$L = \frac{2}{K}$，中央値$\mu_k = -1 + L\left(k+\frac{1}{2}\right)$と定義する．そして，画像内のピクセルがどの画素値レベル（区間）に属するのかを Deephist のように次式で定義する．

$$P_I(k) = \int_{\mu_k-\frac{L}{2}}^{\mu_k+\frac{L}{2}} f_I(g)dg = \frac{1}{N}\sum_{x\in\Omega}\Pi_k(I(x)) \tag{3}$$

ここで，

$$\Pi_k(z) \triangleq \sigma\left(\frac{I(x)-\mu_k+\frac{L}{2}}{B}\right) - \sigma\left(\frac{I(x)-\mu_k-\frac{L}{2}}{B}\right) \tag{4}$$

とする．最終的には式（3）の右辺を用いて生成画像の色分布ヒストグラムを推定する．

次に，色分布ヒストグラムの損失関数について述べる．式（3）によって推定された生成画像X_{fake}，Y_{fake}の色分布ヒストグラムは以下のように表せる．

$$\boldsymbol{h}_{\mathrm{fake}} = \{P_I(k)\}_{k=0}^{K-1} \tag{5}$$

色情報制約の損失関数$\mathcal{L}_{\mathrm{hist}}$は生成画像の色分布ヒストグラム$\boldsymbol{h}_{\mathrm{fake}}$とリファレンス画像の色分布ヒストグラム$\boldsymbol{h}_{\mathrm{ref}}$を用いて，以下のように表せる．

$$\mathcal{L}_{\mathrm{hist}} = \left\|\boldsymbol{h}^{\mathrm{R}}_{\mathrm{fake}} - \boldsymbol{h}^{\mathrm{R}}_{\mathrm{ref}}\right\|_2 + \left\|\boldsymbol{h}^{\mathrm{G}}_{\mathrm{fake}} - \boldsymbol{h}^{\mathrm{G}}_{\mathrm{ref}}\right\|_2 + \left\|\boldsymbol{h}^{\mathrm{B}}_{\mathrm{fake}} - \boldsymbol{h}^{\mathrm{B}}_{\mathrm{ref}}\right\|_2 \tag{6}$$

式（6）は各 RGB において，生成画像の色分布ヒストグラムとリファレンス画像の色分布ヒストグラムの L2 損失を用いる．

3.3 色情報を与えた生成器のネットワーク構造

生成器 : G，F のネットワーク構造を図 2 に示す．入力画像は RGB 画像であり，生成器は encoder-decoder 構造を通して生成画像を出力する．また，生成器はリファレンス画像から背景領域を除いた色分布ヒストグラムを算出し，全結合層によって抽出された色情報の特徴量を中間層に取り入れる．3. 2 節でも述べたが，図 2 に示すように，リファレンス画像の色分布ヒストグラムと，カーネル密度推定を用いて得られた生成画像の色分布ヒストグラムに色情報制約として L2 損失を導入する．

Vol.2022-CVIM-229 No.34
2022/3/11

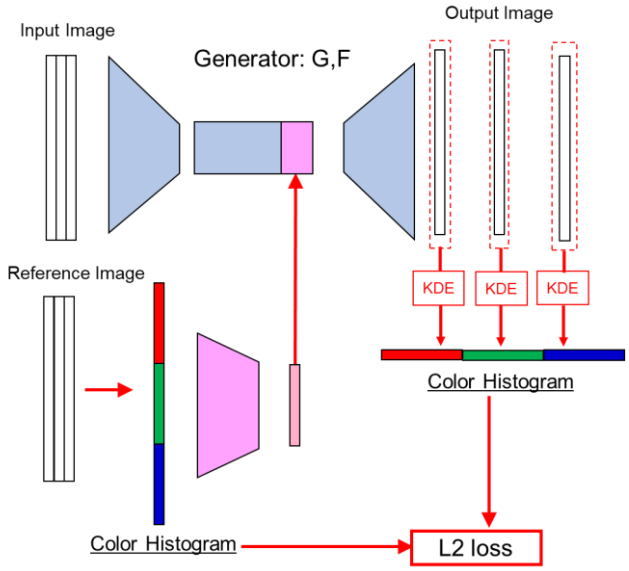

図 2　Generator のネットワーク構造

3.4 ネットワーク全体の損失関数

ネットワーク全体の損失関数$\mathcal{L}$は以下のように 4 つの項で構成される．

$$\mathcal{L} = \lambda_{\text{gan}}\mathcal{L}_{\text{gan}} + \lambda_{\text{cycle}}\mathcal{L}_{\text{cycle}} + \lambda_{\text{idt}}\mathcal{L}_{\text{idt}} + \lambda_{\text{hist}}\mathcal{L}_{\text{hist}} \quad (7)$$

$\mathcal{L}_{\text{gan}}$は敵対的損失，$\mathcal{L}_{\text{cycle}}$は一貫性損失，$\mathcal{L}_{\text{idt}}$は同一性損失，$\mathcal{L}_{\text{hist}}$は色情報損失を表す．

敵対的損失$\mathcal{L}_{\text{gan}}$は GAN で用いられ，以下のように表される．

$$\mathcal{L}_{\text{gan}} = \mathcal{L}_{\text{adv}}(G, D_{\text{Y}}, X, Y) + \mathcal{L}_{\text{adv}}(F, D_{\text{X}}, Y, X) \quad (8)$$

$$\mathcal{L}_{\text{adv}}(G, D_{\text{Y}}, X, Y) = \mathbb{E}_{y\sim p_{data}(y)}[\log D_{\text{Y}}(y)] + \mathbb{E}_{x\sim p_{data}(x)}[\log(1 - D_{\text{Y}}(G(x)))] \quad (9)$$

ここで，xはドメインXのデータ$(x \in X)$，yはドメインYのデータ$(y \in Y)$を表す．また，データ分布のサンプリングを$x\sim p_{\text{data}}(x)$，$y\sim p_{\text{data}}(y)$として表す．

一貫性損失$\mathcal{L}_{\text{cycle}}$，同一性損失$\mathcal{L}_{\text{idt}}$は Cycle GAN で用いられ，以下のように表される．

$$\mathcal{L}_{\text{cycle}} = \mathbb{E}_{x\sim p_{\text{data}}(x)}\left[\left\|F(G(x)) - x\right\|_1\right] + \mathbb{E}_{y\sim p_{\text{data}}(y)}\left[\left\|G(F(y)) - y\right\|_1\right] \quad (10)$$

$$\mathcal{L}_{\text{idt}} = \mathbb{E}_{x\sim p_{\text{data}}(x)}[\|F(x) - x\|_1] + \mathbb{E}_{y\sim p_{\text{data}}(y)}[\|G(y) - y\|_1] \quad (11)$$

一貫性損失$\mathcal{L}_{\text{cycle}}$は形状を保存するように働き，同一性損失$\mathcal{L}_{\text{idt}}$は不要な変換を行わないようなペナルティとして働くことが知られている．

色情報損失$\mathcal{L}_{\text{hist}}$は式（6）で示した通りである．

4. 検証実験

本手法はアメリカの国立衛生研究所（National Institutes of Health）から支援を受ける OAI（OsteoArthritis Initiative）が提供する膝の MRI データセットと，Kaggle が提供する Fashion Product Images Dataset を用いて有効性を検証する．

OAI は臨床データ，患者の報告結果，生物試料分析，定量的画像分析, X 線写真（X-ray）および磁気共鳴画像（MRI）などのデータが含まれるアーカイブを提供する．このアーカイブの膝の MRI データからシミュレータ画像 466 枚を作成した．また，Fashion Product Images Dataset から膝領域を切り抜いた膝の外観画像 505 枚を作成した．学習にはシミュレータ画像 438 枚，膝の外観画像 477 枚を用いて，評価にはシミュレータ画像 28 枚，外観画像 28 枚を用いる．ここで，評価において総当たりの変換$(28 \times 28 = 784)$を行うことによって，より多様な検証を行った．ここで，画像の解像度は256×256としている．

4.1 評価方法

生成画像がリファレンス画像の色を反映しているかの評価を行うために，L*a*b*色空間における色度差を客観評価として用いた．L*a*b*色空間は人間の色に関する知覚に近いと言われている．色度差は以下のように L*a*b*色空間における画素平均のユークリッド距離を表す．

$$\Delta E = \sqrt{(\Delta L^*)^2 + (\Delta a^*)^2 + (\Delta b^*)^2} \quad (12)$$

色度差ΔEは工業製品の色度差を測色する際に用いられる．色度差の許容範囲は表 1 の日本工業規格に示されており，製品の色度差の許容範囲は分野によって異なる．本研究では，この表 1 を元に評価を行う．

表 1　日本工業規格で示される色度差表

	色度差ΔEの範囲	知覚される色差の範囲
A 級許容差	1．6~3．2	色の隣接比較では，ほとんど気付かれない色差レベル．一般的には同じ色だと思われているレベル．
B 級許容差	3．2~6．5	印象レベルでは同じ色として扱える範囲．塗料業界やプラスチック業界では色違いでクレームになることがある．
C 級許容差	6．5~13．0	JIS 標準色票, マンセル色票などの 1 歩度に相当する色差．
D 級許容差	13．0~25．0	細分化された系統色名で区別ができる程度の色の差で，この程度を超えると別の色名のイメージになる．

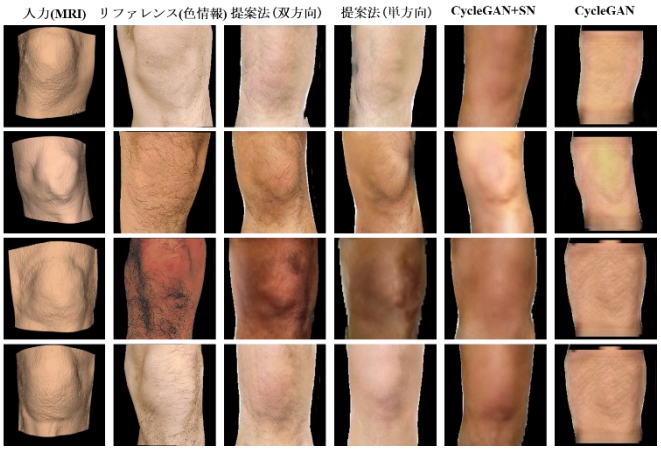

図 3 提案法と CycleGAN の生成画像

評価するモデルは CycleGAN と提案法を対象に検証を行う. CycleGAN に関しては, 従来の CycleGAN と SN(Spectral Normalization)を付与した 2 つのモデルを学習し，検証を行う．一方，提案法に関しては，双方向の生成器G，Fに色情報を導入したモデルと，単方向の生成器Gに色情報を導入したモデルを学習し，検証を行う．

4.2 検証結果

3D モデルのシミュレータ画像群･膝の外観画像群のドメイン適応タスクにおいて，先に述べたデータセットを用いて提案手法や CycleGAN などの複数のモデルで学習を行い，評価を行った．その生成画像を図 3 に示す．

図 3 の左から一列目は生成器Gへの入力画像（シミュレータ画像)，左から二列目は提案法における生成器Gの中間層に与える色情報のリファレンス画像，左から三列目は提案法（双方向）の生成画像，左から四列目は提案法（単方向）の生成画像，左から五列目は従来の CycleGAN に SN を導入した生成画像，左から 6 列目は従来の CycleGAN の生成画像を表す．図 3 の生成画像より，CycleGAN の生成画像は入力の形状を若干ながら担保しているが，リファレンス画像の色情報を用いていないため，生成画像の色は学習したデータの色分布からランダムに生成され，特定の色を生成できていないことが分かる．また，SN を導入した CycleGAN の生成画像も形状に関しては従来よりも担保できているが，特定の色を再現できていない．一方，提案法では，単方向・双方向共に入力の形状を担保し，リファレンス画像の色を再現できていることが分かる．

表 2 は各モデルの生成画像に対する L*a*b*色空間で定義される色度差ΔEを示す．表 2 の色度差ΔEは評価データセットの平均値を表す．表 1 の日本工業規格で表される色度差表より，CycleGAN の生成画像の色度差ΔEは D 級許容差であり，提案法の生成画像の色度差ΔEは単方向・双方向共に B 級許容差であることが分かる．さらに，双方向の方が単方向よりも色度差ΔEが小さくなっており，リファレンス画像の色合いをより反映していることが分かる．双方向が単方向よりもリファレンス画像の色の再現を向上させた理由に関して，逆方向の生成器Fの中間層にも色情報を与えることによって，生成器Fの色情報と形状情報抽出の役割分担が可能になり，生成器Fの入力画像から

表 2 生成画像とリファレンス画像との色度差ΔE

	色度差 ΔE
CycleGAN	14．1
CycleGAN(SN あり)	13．7
提案法(単方向)	5．0
提案法(双方向)	4．6

の形状抽出の効果が強くなったと考える．この効果は一貫性損失により，間接的に生成器Gの入力画像からの形状抽出の効果も大きくし，生成器Gの色情報と形状情報抽出の役割分担をも向上させたと推察する．

また，提案法（双方向）に関して，入力画像とリファレンス画像の様々な組み合わせによる生成画像を図 4 に示す．一番左の列は生成器Gへの入力画像，一番上の行はリファレンス画像を表す．この結果も同様に，提案法の生成画像は入力画像の形状を維持しながら，様々な色分布を持つリファレンス画像の色を再現できることが分かる．この結果は CycleGAN への色分布制約の適用が機能していることを示している．

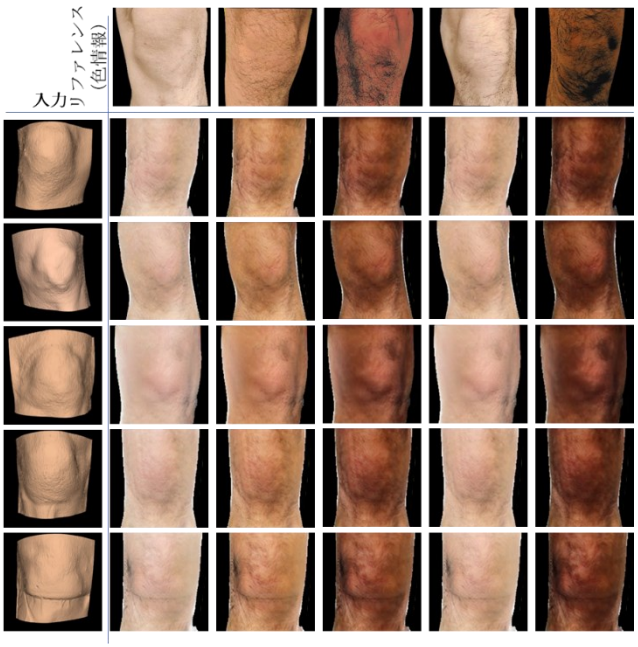

図 4 様々なリファレンスによる提案法の生成画像

5. おわりに

本論文では，膝の 3D シミュレータ画像と外観画像の画像変換において，特定の人物の皮膚色・明るさに合わせるドメイン内の色分布制御が可能なドメイン適応手法を提案した．CycleGAN の特徴である少数かつペアでないデータを用いるドメイン変換にカーネルに密度推定を用いた色情報制約を導入することによって, 3D シミュレータ画像の形状を維持しながら，様々な皮膚色・明るさを指定して再現することができた. しかしながら, 学習過程が進むにつれ，色の再現度が上がるが，形状の再限度が落ちる現象が起きた．この現象は過学習により，形状の陰影がなくなって色分布が単純になり, 色の再現が簡単になったと推察できる．つまり，形状の再現と色の再現はトレードオフの関係になっている．今後の課題として，このトレードオフの関係を打開するような手法の調査・提案を行う必要がある．

謝辞

本論文執筆で使用された MRI データは, NIMH データアーカイブ（NDA）内のデータリポジトリである Osteoarthritis Initiative（OAI）のデータセットから取得されました． OAI は，国立精神衛生研究所と国立関節炎・筋骨格・皮膚疾患研究所（NIAMS）によって作成された共同インフォマティクスシステムであり, バイオマーカーの特定, 科学的調査, および OA 医薬品開発のペースを速めるための世界的なリソースを提供します． データセット識別子：2343

参考文献

[1] I. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, and Y. Bengio. Generative adversarial nets. Advances in Neural Information Processing Systems (NeurIPS), pages 2672–2680, 2014.

[2] J. Zhu, T. Park, P. Isola, and A. A. Efros. Unpaired image-to-image translation using cycle-consistent adversarial networks. International Conference on Computer Vision (ICCV), to appear, 2017.

[3] P. Isola, J.-Y. Zhu, T. Zhou, and A. A. Efros. Image-to-image translation with conditional adversarial networks. International Conference on Computer Vision and Pattern Recognition (CVPR), 2017.

[4] Mor Avi-Aharon, Assaf Arbelle, and Tammy Riklin Raviv. Deephist: Differentiable joint and color histogram layers for image-to-image translation. arXiv preprint arXiv:2005.03995, 2020.

[5] Yunjey Choi, Youngjung Uh, Jaejun Yoo, and Jung-Woo Ha. Stargan v2: Diverse image synthesis for multiple domains. International Conference on Computer Vision and Pattern Recognition (CVPR), pages 8188–8197, 2020.

[6] Choi, Y., Choi, M., Kim, M., Ha, J.W., Kim, S., Choo, J.: Stargan: Unified generative adversarial networks for multi-domain image-to-image translation. International Conference on Computer Vision and Pattern Recognition (CVPR), 2018.

[7] T. Karras, S. Laine, and T. Aila. A style-based generator architecture for generative adversarial networks. International Conference on Computer Vision and Pattern Recognition (CVPR), 2019.

[8] Tero Karras, Samuli Laine, Miika Aittala, Janne Hellsten, Jaakko Lehtinen, and Timo Aila. Analyzing and improving the image quality of stylegan. International Conference on Computer Vision and Pattern Recognition (CVPR), 2020.

[9] A. Radford, L. Metz, and S. Chintala. Unsupervised representation learning with deep convolutional generative adversarial networks. arXiv preprint arXiv:1511.06434, 2015.

[10] A. Brock, J. Donahue, and K. Simonyan. Large scale GAN training for high fidelity natural image synthesis. CoRR, abs/1809.11096, 2018.

[11] Wasserstein GAN, by Martin Arjovsky et al.,2017,https://arxiv.org/pdf/1701.07875.pdf

[12] T. Miyato, T. Kataoka, M. Koyama, and Y. Yoshida. Spectral normalization for generative adversarial networks. CoRR, abs/1802.05957, 2018.