The 32nd Annual Conference of the Japanese Society for Artificial Intelligence, 2018

CycleGAN を用いた感情スタイル転送

Emotional Style Transfer using CycleGAN

大田 和寛 *1
Kazuhiro Ota

*1株式会社サイバーエージェント
CyberAgent, Inc.

Our emotions or impressions to the images largely depend on the visual features of each image. If people in a photography are smiling, we might perceive a happy impression. On the other hand, if violent aliens are drawn, we may feel scary. We consider an emotional style transfer model that transforms an impression of the given image to the desired one. We utilize adversarial training of pixelwise transformation to learn a model with Behance Artistic Media dataset and report initial results.

1. はじめに

近年，様々なストックフォトサービスの台頭により容易に高精細な画像を手に入れることが可能となった．画像に付与されているタグで検索するだけでそのタグに該当する画像を収集できるため，バナー広告のクリエイティブ制作現場などでも素材候補として頻繁に利用されている．しかしながら，ストックフォトサービスにおけるタグデータは基本的にユーザが付与するものであり，写真の被写体や画像の持つ印象に大きく依存するため，しばしば検索のノイズとなり易い [Yang 17]．Wilber らはこのような画像とメタデータとのミスマッチ問題に対応するために Behance Artistic Media（BAM）*1 というデータセットを提案している [Wilber 17]．これは画像の被写体，表現媒体，感情における全 20 種の属性について，クラウドソーシングによりそれぞれの属性に対するスコア付けを行いメタデータとして添付したデータセットとなっており，コンピュータビジョンによる写真に留まらない画像表現の理解・学習の促進を期待するものである．

一方で，コンピュータビジョンにおける Deep Learning の応用事例の中でも画像表現に着目した興味深い研究が報告されている．Gatys らはコンテンツ画像とスタイル画像を入力としてコンテンツ画像に対しスタイル画像の画風のみを転送するといった画風スタイル変換手法を提案している [Gatys 16]．Goodfellow らはノイズから画像を生成する生成器ネットワークとその生成画像に対し本物・偽物の判断を行う判別器ネットワークとを連結させて互いに競い合うように学習することで本物に近い画像を自動生成する Generative Adversarial Network（GAN）と呼ばれるネットワーク構造を提案している [Goodfellow 14]．Isola らが提案した手法である pix2pix は画像ペアごとの変換規則およびそれぞれのドメイン間で固有の損失関数を取得することで手書きエッジから写実的な物体を生成するような所謂 Image-to-Image と呼ばれる変換を行っている [Isola 17]．また，Zhu らはさらにこの pix2pix から学習データのペア制約を取り除くことにより，教師なし学習の枠組みで Image-to-Image を行う CycleGAN を提案している [Zhu 17]．

本稿では，BAM に含まれる感情属性のスコアをもとにデータセットを “happy”，“scary” に分割し，感情ドメインにおける Image-to-Image を CycleGAN にて行うことで感情表現の相互転送手法を提案する（図 1）．画像データから感情表現を抽出することができれば，幸福や恐怖といった画像が有する印象をもとに素材を加工することが可能となる．これは例えばバナー広告のデザイン元の素材画像に対して感情を付与できることを意味している．ストックフォトサービスの画像におけるタグやフリーワードなどのメタデータの信頼度を強固にできるなどの応用が考えられ，バナー広告の制作現場に対する支援の一手となり，これまで手作業で行っていた印象に基づく画像操作の指標となり得るものである．



図 1: 本稿では BAM に含まれる “happy” と “scary” という 2 つの感情ドメインの画像群に対し CycleGAN を用いた双方向の感情スタイル転送手法を提案する，笑顔の女性をまるでホラー映画のキャラクターのように変換し，骸骨や物寂しい風景画をカラフルに彩色している．

2. 提案手法

2.1 BAM

BAM に含まれているメタデータの中で本稿において重要なものを以下に簡潔にまとめておく．詳細は原著論文 [Wilber 17] を参照されたい．

連絡先: 大田和寛，株式会社サイバーエージェント，ota_kazuhiro@cyberagent.co.jp

*1 https://bam-dataset.org/

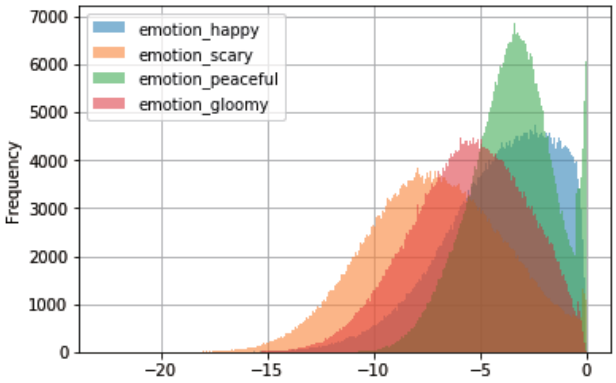


図 2: BAM における感情ドメインの属性 4 種 “emotion_happy”，“emotion_scary”，“emotion_peaceful”，“emotion_gloomy”のスコア分布．

属性
画像の被写体（“content_people”，“content_car”など），表現媒体（“media_pen-ink”，“media_comic”など），感情（“emotion_happy”，“emotion_scary”など）によって分類された全 20 種類の属性名である．

ラベル
クラウドソーシングとラベル付アルゴリズムとの組み合わせである Human-in-the-Loop 手法 [Yu 15] により，画像が上記 20 種のいずれかの属性として妥当性を持つか否かの判断結果が “positive”，“unsure”，“negative”の 3 値で与えられている．

スコア
ラベル同様 Human-in-the-Loop 手法によって得られた値だが，画像 1 件に対し全属性分の信頼度が負値の連続値で格納されている．図 2 に感情ドメインの属性におけるスコアの分布を示す．

図 3 に BAM における感情ドメインの属性である “emotion_happy”と “emotion_scary”におけるスコア上位の画像群を幾つか示す．いずれも双方の感情表現が画像によく現れており，スコアでのソート結果が画像特徴の理解を促進するものであることが伺える．

2.2 CycleGAN

ある画像集合 X および Y があったとき，CycleGAN は X，Y をそれぞれ互いに変換する $G^\star : X \to Y$，$F^\star : Y \to X$ を学習する．その際の損失関数は以下のように設計されており，

$$\begin{aligned}\mathcal{L}(G, F, D_X, D_Y) &= \mathcal{L}_{\mathrm{GAN}}(G, D_Y, X, Y) \\ &+ \mathcal{L}_{\mathrm{GAN}}(F, D_X, X, Y) \\ &+ \mathcal{L}_{\mathrm{cyc}}(G, F), \end{aligned} \quad (1)$$

これを以下の mini-max 式で解くことで得られる変換が $G^\star$，$F^\star$ である．

$$G^\star, F^\star = \arg\min_{G,F} \max_{D_X, D_Y} \mathcal{L}(G, F, D_X, D_Y). \quad (2)$$

ここで D_X，D_Y はそれぞれ $F : Y \to X$，$G : X \to Y$ によって変換された画像に対する GAN 判別器 [Goodfellow 14] であり，式 (1) の第 1 項，第 2 項に示す $\mathcal{L}_{\mathrm{GAN}}$ によって F，G に

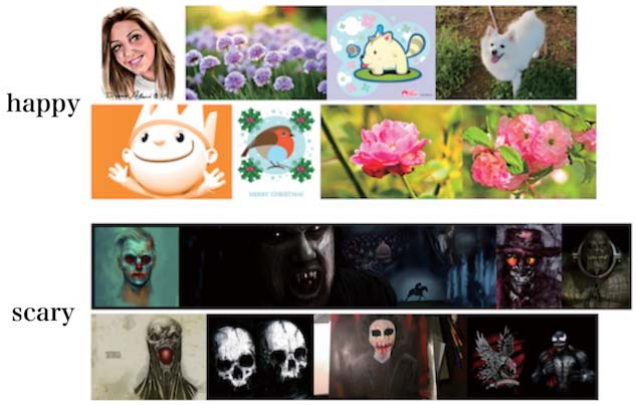


図 3: BAM における “emotion_happy”属性と “emotion_scary”属性のスコア上位に属する画像の例．“emotion_happy”には動物や色鮮やかな草花が多く含まれ，“emotion_scary”にはダークな色合いで凶暴性をもったキャラクターが見受けられる．

よる変換結果に対する敵対的損失を計上する．さらに式 (1) の第 3 項に示す $\mathcal{L}_{\mathrm{cyc}}$ は下記のように定式化されており，

$$\begin{aligned}\mathcal{L}_{\mathrm{cyc}}(G, F) &= \mathbb{E}_{x \sim p_{\mathrm{data}}(x)}[\|F(G(x)) - x\|_1] \\ &+ \mathbb{E}_{y \sim p_{\mathrm{data}}(y)}[\|G(F(y)) - y\|_1], \end{aligned} \quad (3)$$

X，Y 両ドメインの画像に対し F，G を交互に作用させた際のオリジナルデータとの差異の期待値を復元損失として計上するものである [Zhu 17]．近年の応用事例としては情報隠蔽への応用 [Chu 17] や，人間の顔をうまく交換した提案 [Jin 17] も報告されており，非ペア制約条件下におけるドメイン間のスタイル変換に対し印象的な結果をもたらしている．

2.3 概要

BAM データセットの属性種別のうち “emotion_happy”および “emotion_scary”のスコアでの上位 1,000 件分の画像群をそれぞれ “happy”ドメイン，“scary”ドメインとし，CycleGAN によりこのドメイン間で画風スタイルの変換学習を行う．図 4 に概要図，図 5 および図 6 に両ドメインの学習データにおける属性とラベルのペアの分布図を示す．“happy”ドメインを X，“scary”ドメインを Y として学習データと考えた時，$G : X \to Y$，$F : Y \to X$ を CycleGAN によって最適化することに相当し，それぞれのドメインが相互に教師ドメインの役割を果たすため，BAM のように明確な教師ラベルデータが付与されていないデータ群に対し確率的な学習効果を発揮すると考えることができる．

3. 実験

3.1 概要

画像のサイズは全て 286 × 286 ピクセルにリサイズし，最初の 100 エポックは学習係数 0.0002，次の 100 エポックは学習係数 0 まで線形減衰させる形で合計 200 エポックの学習を，PyTorch による実装 *2 を利用し CUDA を介した NVIDIA GTX 1080 Ti 上で実施した．

*2 https://github.com/junyanz/pytorch-CycleGAN-and-pix2pix

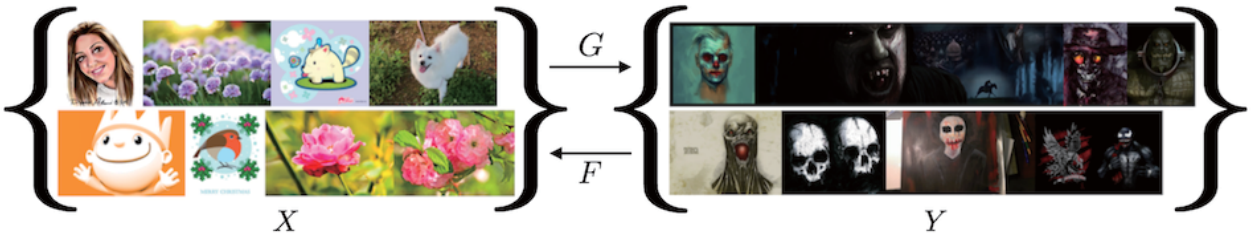


図 4: CycleGAN よる感情ドメイン間での画風変換学習の概要図．BAM から抽出した “happy” ドメイン X と “scary” ドメイン Y の間のギャップを相互変換する関係 $G : X \to Y$ と $F : Y \to X$ をペア制約のない教師なし手法で学習する．

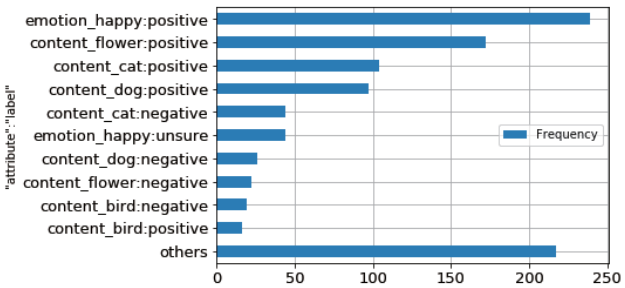


図 5: “happy” ドメイン学習データに含まれる属性値とラベルのペアの内訳．“emotion_happy:positive”，“content_flower:positive”，“content_cat:positive” などが上位を占めている．

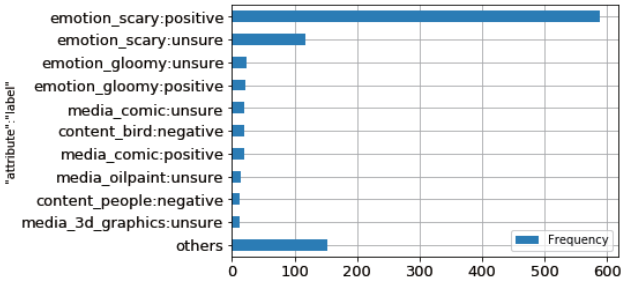


図 6: “scary” ドメイン学習データに含まれる属性値とラベルのペアの内訳．“emotion_scary:positive”，“emotion_scary:unsure”，“emotion_gloomy:positive” などが上位を占めている．

3.2 結果考察

“emotion_happy” および “emotion_scary” のスコアでの上位 1,001〜1,200 件の画像をそれぞれ学習された変換 $G^\star$，$F^\star$ にかけ，得られた画風スタイル変換結果に対し官能的評価を行った．

結果例を図 1 及び図 7 に示す．“happy” →“scary” に関してはカラフルで明るい色味を持った画像が赤黒色を保存しつつエッジをより際立たせるような形での減色がなされ，“scary” →“happy” に関してはモノクロや線画の画像を人間や植物の色に近い形で色付けされエッジを平滑化するような効果が見受けられる結果となった．これは図 5 と図 6 に示した学習データにおける属性値とラベルのペアの分布がうまく反映された結果だと考えられ，CycleGAN による明示的なラベルデータの準備を行うことのない感情表現の転送が達成できたと言える．また，今回の感情表現転送においてあまり視覚的な効果が見受けられなかった例を図 8 に幾つか示す．それぞれのドメインが片方のドメインの特徴を有している場合において期待する効果が得られなかったと考えられる．

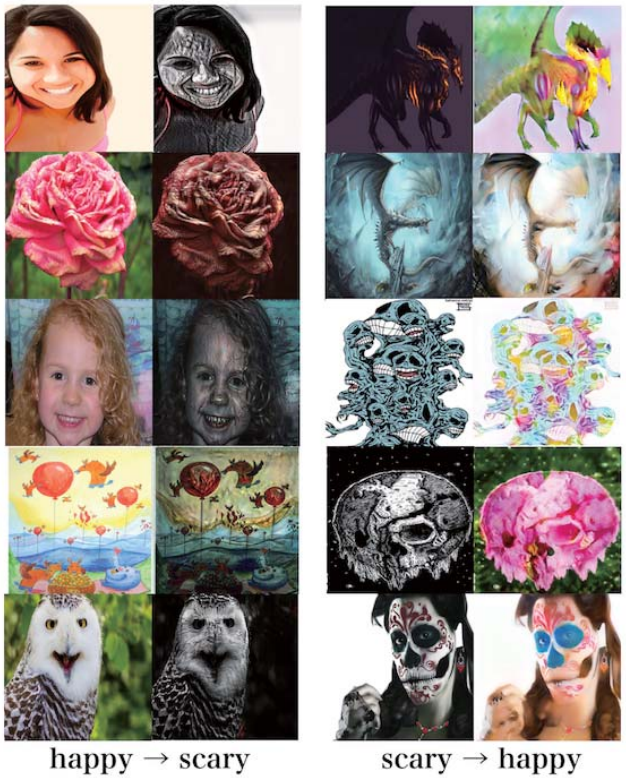


図 7: BAM 感情スタイル変換実験における結果例．“happy” → “scary” においては赤黒色を保存しつつ明度・彩度を下げエッジを際だたせるシャープネス処理がかかった効果が学習された．一方，“scary” → “happy” においては人間の肌色や髪色を考慮しつつ自動彩色し平滑化処理がかかりエッジがぼかされるような効果が確認できる．

4. まとめ

本稿では BAM データセットがもつ属性における 4 種類の感情属性のうち “happy” および “scary” の 2 ドメイン間に存在する画風表現の違いを把握し，CycleGAN を用いた実験によりそれぞれのドメイン間での画風スタイルの転送を行った．笑顔の人間の顔に痣が現れたり，異形のエイリアンが草花のように自動彩色されたりといった画風スタイル変換が行われ，感情ドメイン間における表現の違いがそれぞれの画像群に対して視覚的な変化を与えることが確認できた．これはバナー広告



図 8: BAM 感情スタイル変換実験において視覚的な効果が見受けられなかった結果．“happy”の画像に黒色が多かったり，“scary”の画像がカラフルであったりと，スタイル転送元の画像が片方のドメインの特徴を既にある程度有していることが伺える．

のクリエイティブに利用される素材のような高精細画像に対しても有効に活用できることが見込まれるため，素材データのバリエーションの拡張に繋がると考えられる．しかしながら，今回は 2 種類の感情のみの CycleGAN での画風スタイル転送結果しか得られていないため，定量的評価は今後の課題である．また，画風の転送は行えたものの，“happy”ドメインの女性の笑顔が笑顔のままであったり，“scary”ドメインの骸骨が骸骨のままであったりとコンテンツへの考慮の不足といった問題もある．今後，損失関数の設計など，改善点を見つけ何らかの対応を進めていきたい．

参考文献

[Chu 17] Chu, C., Zhmoginov, A., and Sandler, M.: CycleGAN, a Master of Steganography, in *NIPS* (2017)

[Gatys 16] Gatys, L. A., Ecker, A. S., and Bethge, M.: Image Style Transfer Using Convolutional Neural Networks, in *CVPR* (2016)

[Goodfellow 14] Goodfellow, I., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., Courville, A., and Bengio, Y.: Generative Adversarial Nets, in *NIPS* (2014)

[Isola 17] Isola, P., Zhu, J.-Y., Zhou, T., and Efros, A. A.: Image-to-Image Translation with Conditional Adversarial Networks, in *CVPR* (2017)

[Jin 17] Jin, X., Qi, Y., and Wu, S.: CycleGAN Face-off, *arXiv*, Vol. 1712.03451, (2017)

[Wilber 17] Wilber, M. J., Fang, C., Jin, H., Hertzmann, A., Collomosse, J., and Belongie, S.: BAM! The Behance Artistic Media Dataset for Recognition Beyond Photography, in *ICCV* (2017)

[Yang 17] Yang, F., Kale, A., Bubnov, Y., Stein, L., Wang, Q., Kiapour, M. H., and Piramuthu, R.: Visual Search at eBay, in *KDD* (2017)

[Yu 15] Yu, F., Zhang, Y., Song, S., Seff, A., and Xiao, J.: LSUN: Construction of a Large-scale Image Dataset using Deep Learning with Humans in the Loop, *arXiv*, Vol. 1506.03365, (2015)

[Zhu 17] Zhu, J.-Y., Park, T., Isola, P., and Efros, A. A.: Unpaired Image-to-Image Translation using Cycle-Consistent Adversarial Networkss, in *ICCV* (2017)