

特別研究報告題目

GAN を用いた被写体ブレの補正

Correction of object motion blur using GAN

指導教員 出口利憲 教授

岐阜工業高等専門学校 専攻科 先端融合開発専攻

2021Y06 氏名板津敬佑

令和6年(2024年)2月1日提出

Abstract

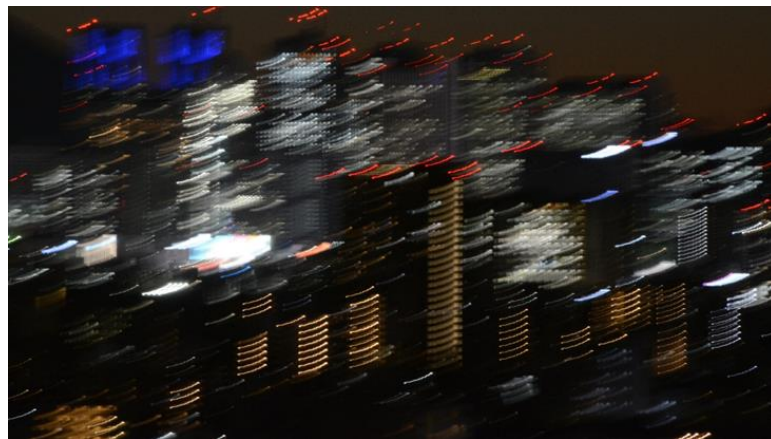
Image deblurring technology has advanced significantly, and many methods have been proposed for deblurring. However, most of them are designed to prevent camera shake, and there are few methods that are effective against object motion blur. Therefore, in this research, we aim to remove subject blur from images by using ESRGAN, a method in the field of super-resolution that has excellent perceptual quality. To perform pair learning, we prepare a dataset by adding artificial blur to the images. Then, in order to make the input and output image sizes the same, we removed the upsampling layer and performed learning. Although we were not able to obtain satisfactory results, we confirmed that ESRGAN, a method in the field of super-resolution, can be used to correct object motion blur.

目次

第1章	序論	2
第2章	関連研究	2
	2.1 敵対的生成ネットワーク	2
	2.2 従来手法	2
第3章	提案手法	4
	3.1 概要	4
	3.2 提案手法	4
第4章	結果と考察	7
	4.1 データセット	7
	4.2 実験方法	7
	4.3 結果	7
	4.4 考察	10
第5章	結論	12
	5.1 結論	12
	5.2 今後の課題	12
参考文献		13
謝辞		14

第1章 序論

近年、画像のブレ除去技術は大幅に進歩しており、またブレの除去のための多数の方法が提案されている。しかし、それらのほとんどはカメラの動きによって生じるブレである手ブレ用に設計されており、被写体ブレを削除するために提案された手法は少ない。被写体の動きとカメラとの相対的な動きによって生じる被写体ブレは、被写体の様々な動きによってブレが生じるため画像全体への均一的なブレ除去方法は効果的に適用することは出来ない。そこで画像内の特定の領域にのみブレ除去を行う手法を提案する。手ブレした画像と被写体ブレした画像の例を Fig.1 に示す。



(a) Camera shake image.



(b) Object motion blur image.

Fig.1 Camera shake image and object motion blur image.

第 2 章 関連研究

2.1 敵対的生成ネットワーク

近年、画像処理の分野におけるディープラーニングの技術は日々進化しており、応用研究が進められたり、ビジネスに適用されたりしている。中でも最近注目されている技術の一つに「敵対的生成ネットワーク」(以下 GAN) がある。GAN は生成モデルの一種であり、データから特徴を学習することで、実在しないデータを生成したり、存在するデータの特徴に沿って変換が行える。GAN は教師なし学習の手法として知られており、そのアーキテクチャの柔軟性から幅広い分野への応用が期待できる。GAN は生成器と識別器の二つのニューラルネットワークで構成されており、生成器が生成した画像を識別器で評価し交互に学習していく事で生成モデルを完成させる [1]。

2.2 従来手法

2.2.1 Deblur-GAN

Deblur-GAN は識別器に Relativistic Discriminator を採用しており、これにより既存手法では識別器で Wasserstein 距離を正確に近似するために何回も識別機を反復していたが、Relativistic Discriminator では生成器と識別器の両モデルを 1 ステップで一回ずつ更新するため学習が早いことが特徴となっている。また距離の最小化(Wasserstein 距離)と真贋判定の両方の性質を持ち、本物か偽物か分からない時に今までは全て本物であると出力していたところを、確率は 0.5 と出力する方向に学習させている。実際にこの学習方法は精度の向上に寄与している。この手法は手ブレや光源等によってもたらされる画像全体のノイズの除去を目的として盛んに研究されている手法である [2]。

2.2.2 SRGAN

SRGAN は低解像度の画像から高解像度の画像を出力する超解像に GAN を初めて適用したモデルであり、SRGAN 以前の手法に比べて結果の知覚品質が高く自然であることが特徴である。ESRGAN は SRGAN を前身にもち、SRGAN の課題であ

った学習画像とテスト画像の差が大きいときに生じるノイズを解消するため、ネットワーク構造の畳み込み層から活性化関数に値を入力する間にある正規化する層を取り除く事でノイズを抑制し、計算量を減らすことにも成功している[3][4]。

第3章 提案手法

3.1 概要

本章では、超解像の分野の手法を用いて、被写体ブレが入った画像からブレの無い画像を生成する方法について述べる。

3.2 提案手法

本研究では超解像の手法である ESRGAN[4]のネットワークに改良を加えて、被写体ブレを取り除く生成モデルを作成する。ESRGAN の生成器は画像のピクセルを用いて本物画像と生成画像の Pixel 毎の差から損失を計算する Pixel loss、本物画像と生成画像の特徴抽出した結果を用いて特徴抽出した結果を用いてそれぞれの特徴量同士の平均二乗誤差で損失を計算する Perceptual loss、生成器と識別機の出力結果を用いて、相対的に一方(本物、生成)の画像がもう一方の画像よりも本物である確率を計算する Adversarial loss の3つの損失関数を最小化するように学習する。生成器のネットワークは、畳み込み層(Convolution 層)、23 個の Basic Block (Residual-in-residual Dense Block(以下 RRDB))、畳み込み層、アップサンプリングレイヤー、そして2つの畳み込み層から構成される RRDB は Dense Block の1つであり、その入力をブロックの出力に加算するスキップ接続を備えている。これにより、ネットワークは恒等関数を学習することが可能となり、層を深くしても勾配が消失すること無く効果的な学習が可能となっている。基本的な ESRGAN の生成器の構成を Fig.2 に示す。

なお、本研究では画像のブレに対してのみ補正を行うため、アップサンプリングは行わずに学習を行う。

ESRGAN は SRGAN の Batch Normalization 層 (BN 層) と呼ばれる訓練データの平均と分散から特徴量の正規化を行う部分を削除しており、これにより SRGAN の課題であった訓練データとテストデータが大きく異なるときに生じるノイズを克服している。その構成を Fig.3 に示す。

RRDB は、層数や接続数を増やすことで性能が向上するという予測に基づいて、SRGAN のオリジナルの残差ブロックよりも深く複雑な Residual-in-residual の構造を採用しており、Dense Block を弱い残差で結合している。RRDB を用いたより深いモデルは、例えば画像中の規則的な構造の意味情報をとらえる表現力が高く、生成画

像のテクスチャをさらに向上させることが出来る。モデルの図を Fig.4 に示す。

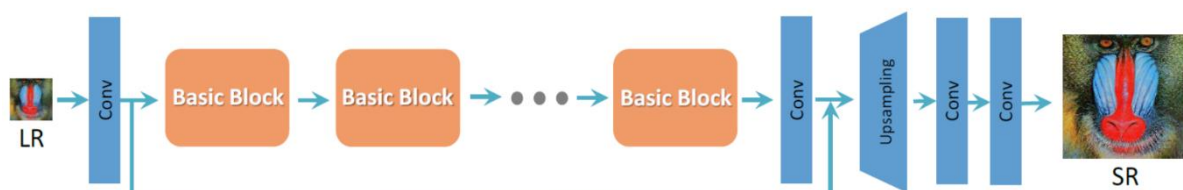


Fig.2 Network structure of the generator[4].

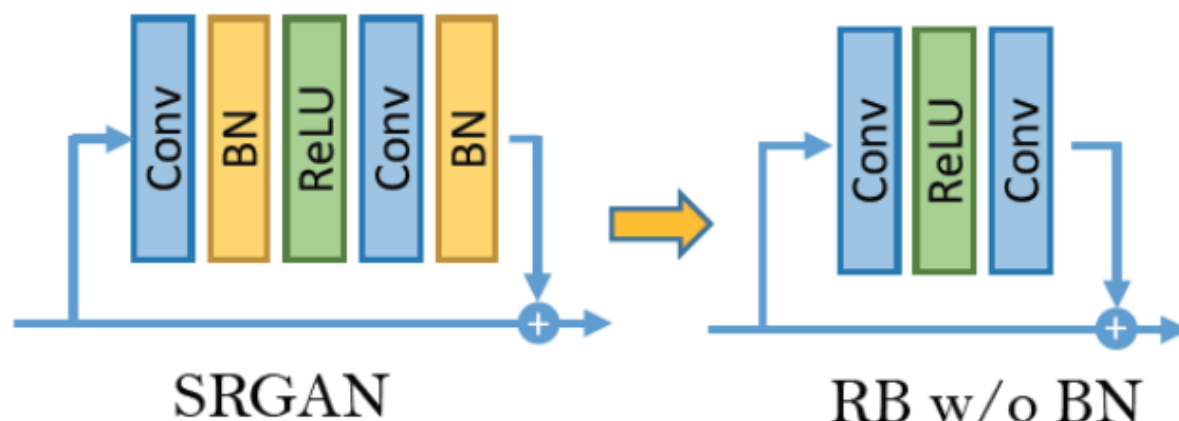


Fig.3 Structure of residual block[4].

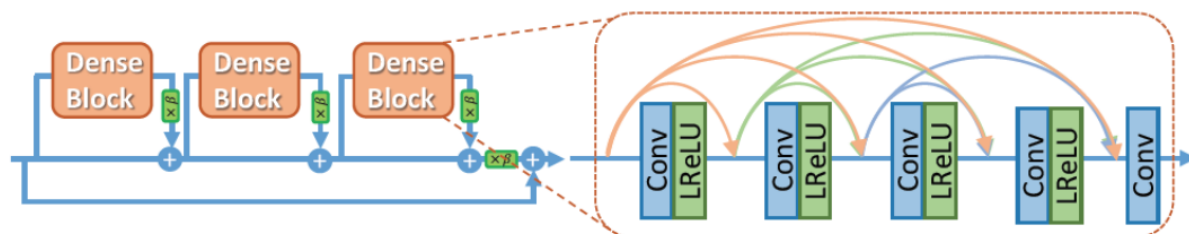


Fig.4 Structure of RRDB[4].

GAN の学習は 2 段階で構成されており、事前学習では本物画像と低解像度画像との間の L1 ノルムの距離を評価する損失関数 Pixel loss を用いる。G を生成器, x_i を低解像度の本物画像、 $G(x_i)$ を生成画像、 y を本物画像とした時、

$$L_1 = E_{x_i} \|G(x_i) - y\|_1$$

となる。この時 $E_x[\cdot]$ はミニバッチ内のすべてのデータの平均を取る演算をする。画素単位の損失を用いた事前学習を行うことで、より自然な超解像画像を生成することが出来る。この事前学習を終えた後、ほかの損失関数も加えて学習を行う。特徴量抽出器 F によって得られた特徴量を用いる Perceptual loss を用いて、L1 ノルム距離が最小するように学習する[3]。

$$L_{percep} = E_{x_i} \|F(G(x_i)) - F(y)\|_1$$

このとき、 $F(F())$ は画像から得られる特徴マップである。次に、生成器と識別機の出力結果を用いて、相対的に一方(本物、生成)の画像がもう一方の画像よりも本物である確率を計算する Adversarial loss は以下の通りになる。

$$L_G^{Ra} = -E_{x_r} [\log(1 - D_{Ra}(x_r, x_f))] - E_{x_f} [\log(D_{Ra}(x_f, x_r))]$$

の時 $D_{Ra}(a, b)$ は a が b より本物であるほど 1 に近づき、本物でないほど 0 に近づく。

$$D_{Ra}(x_f, x_r) = \sigma(C(x_f) - E[C(x_r)])$$

$$D_{Ra}(x_r, x_f) = \sigma(C(x_r) - E[C(x_f)])$$

σ : 活性化関数

x_r, x_f : 本物画像, 生成画像

$C()$: 生成器の出力

$E[]$: 期待値 (ミニバッチの全データの平均)

と定められる。

Adversarial Loss はそれぞれの生成器の出力値の差に対して 2 値公差エントロピー損失を取るよう計算される。生成器のパラメータ更新時は $D_{Ra}(x_r, x_f)$ が 0 になり、 $D_{Ra}(x_f, x_r)$ が 1 になるように学習する。生成器の Adversarial loss を用いることでより詳細なテクスチャの学習ができる。以上 3 つの損失関数を用いて学習し、異なる損失項のバランスをとるための重み係数 λ, η をかけることで算出した値を生成器の Loss として、生成器の学習を行う[4]。

$$L_G = L_{percep} + \lambda L_G^{Ra} + \eta L_1$$

識別機の損失関数においては生成器の Adversarial Loss と同様に以下で定義される[2]。

$$L_D^{Ra} = -E_{x_r} [\log(D_{Ra}(x_r, x_f))] - E_{x_f} [\log(1 - D_{Ra}(x_f, x_r))]$$

前述の生成器とは逆にパラメータの更新時は $D_{Ra}(x_r, x_f)$ が 1 になり、 $D_{Ra}(x_f, x_r)$ が 0 になるように学習する[4]。

第4章 実験結果と考察

4.1 データセット

本来であれば、実際に被写体ブレした画像を用意することが好ましいがそれらを学習用のデータセットに必要な枚数を用意することは困難である。そのため本研究では、ブレのない画像の被写体の画素を一方向にずらし人工的なブレを追加したものをブレ画像とし、ペア学習を行った。また、ブレ画像と本物画像をセットとしそれぞれ60枚ずつで、ブレの大きさを変えて2回実験を行った。データセットの例を Fig.5 に示す。



(a) Big blur (b) Small blur (c) Real image

Fig.5 Learning images.

4.2 実験方法

本研究では、同データセットを変更して2回の学習を行い、2つの生成モデルを作成した。一度目の学習では大きいブレを追加して学習を行った。二度目の学習では一度目と比較して小さいブレを追加して学習を行った。ネットワークモデルは、Kuan-yu Huang氏がGithubにアップロードしたもの[5]を用いる。また、超解像用のネットワークを被写体ブレの補正に流用するためにアップサンプリング層を削除し実験を行った。データのサイズは計算資源を考慮し 128×128 Pixelを入力とした。実際の学習の際は学習用画像からランダムに 32×32 の領域を抜き出して学習を行う。

4.3 結果

それぞれの生成モデルの学習過程における損失を Fig.6 に示す。

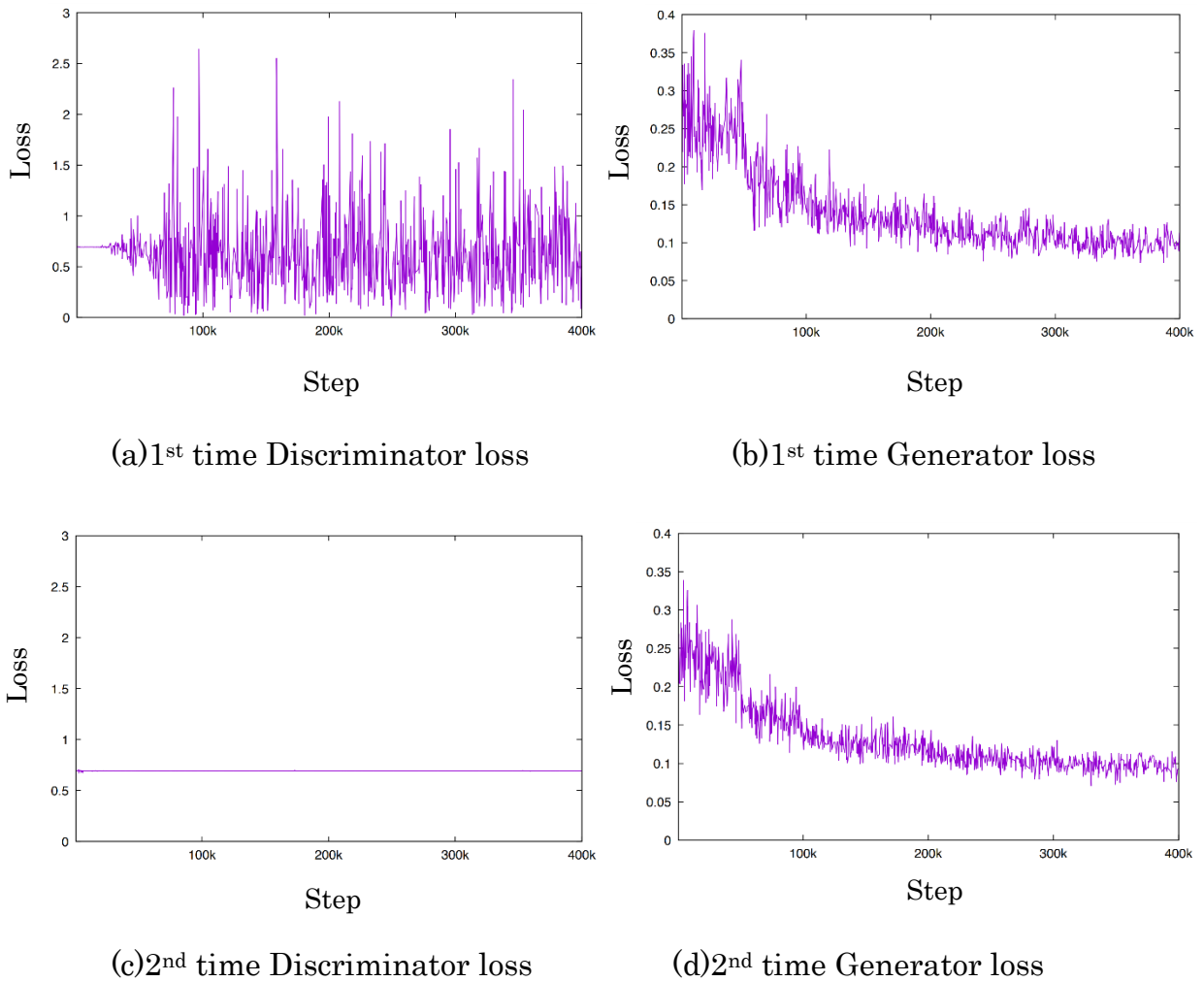
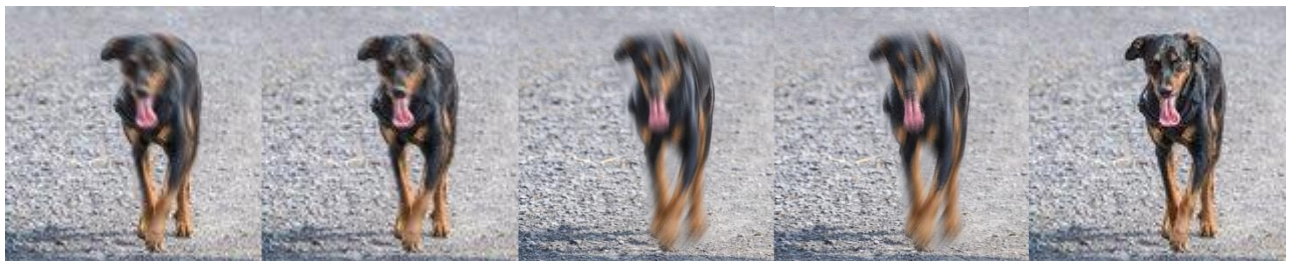


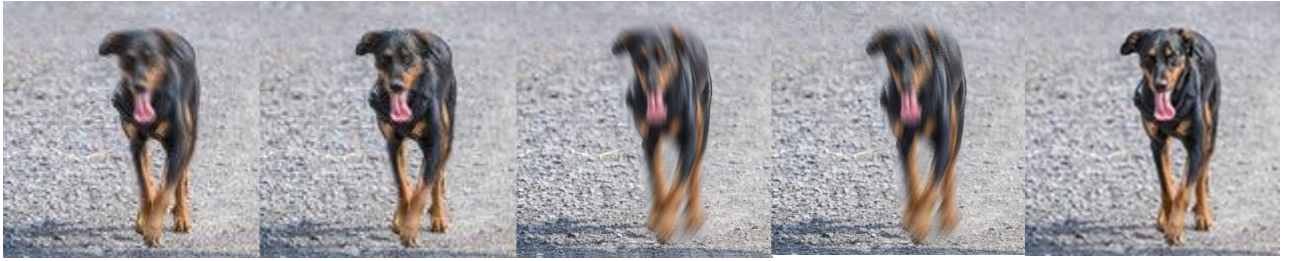
Fig.6 Generator and Discriminator loss

次に1回目に行った、学習にブレの大きい画像を用いたモデルに、新規のブレ画像に異なるブレを追加した2つのテスト画像を入力したテスト結果を Fig.7、2回目の学習モデルから得られた結果を Fig.8 に示す。同様に1回目の学習モデルに、学習に使用したブレ画像を入力したテスト結果を Fig.9、Fig.11、2回目の学習モデルから得られた結果を Fig.10、Fig.12 に示す。



(a) Small blur (b) Small blur output (c) Big blur (d) Big blur output (e) real image

Fig.7 Model trained on highly blurred images(Doberman).



(a) Small blur (b) Small blur output (c) Big blur (d) Big blur output (e) real image

Fig.8 Model trained on slightly blurred images(Doberman).



(a) Small blur (b) Small blur output (c) Big blur (d) Big blur output (e) real image

Fig.9 Model trained on highly blurred images(Bulldog).



(a) Small blur (b) Small blur output (c) Big blur (d) Big blur output (e) real image

Fig.10 Model trained on slightly blurred images(Bulldog).



(a) Small blur (b) Small blur output (c) Big blur (d) Big blur output (e) real image

Fig.11 Model trained on highly blurred images(cat).



(a) Small blur (b) Small blur output (c) Big blur (d) Big blur output (e) real image

Fig.12 Model trained on slightly blurred images(cat).

4.4 考察

Fig.6 より、1 回目の学習では識別器の損失が大きく変動しており、識別機が 100% 見破れるような状態もあれば生成器に騙されている状態もあり理想的な損失であると言える。また生成器の損失が徐々に小さくなっていることから、1 回目の学習は成功していると言える。2 回目の学習では識別器の損失は 0.7 前後を推移し大きく上下していないことが分かる。これは識別器の損失は生成画像と本物画像の差が少ないほど値が小さくなる性質があり、2 回目の学習ではデータセットの本物画像とブレ画像の差が少ないため値が小さくなったと考えられる。しかし、変動していない訳ではなく、生成器の損失は 1 回目と同様なグラフとなっているため学習は成功していると考えられる。以上よりデータセットにおけるブレの大小はモデルの学習の成否においては重要ではないことが分かる。

次に、1 回目に学習を行ったモデルの結果である Fig.7、Fig.9、Fig.11 では、動物の模様には微かなエッジが見られるもののブレが取り除かれているとは言えない結果となった。これは ESRGAN が学習用画像の特定の領域を切り取って学習を行うため、学習用画像のブレが大きい場合に本来学習に必要な画素が領域から漏れてしまうことにより、モデルの性能が落ちたと考えられる。以上の考察より学習用画像のブレを小さくしたデータセットを用いて 2 回目の学習を行った。Fig.9 と Fig.10、Fig.11 と Fig.12 の 2 つの output を比較してみると 2 回目のモデルは 1 回目のモデルより輪郭のぼやけが少なく、目や口といったパーツ単位で注目するとブレを除去できており、性能が向上している。この結果から学習用画像のブレが大き過ぎるとモデルの性能が落ちることが確認できた。

Fig.7、Fig.8 からは学習で使われていないブレ画像はそれ以外の結果と比べてブレの補正がうまく行えていないことが分かる。これは学習に使われていない画像のブレ

と類似するブレが学習用データセットに無いため起こったと考えられる。

第5章 結論

5.1 結論

本研究では、被写体ブレの補正を実現するために超解像の分野の手法であるESRGANを流用する手法を提案した。その結果、被写体ブレした画像からある程度のブレを取り除く事に成功した。また、学習データのブレの入り具合によってモデルの精度に差が出ることも確認できた。

5.2 今後の課題

本研究では被写体ブレした画像からブレを取り除く事が目的であったが、ESRGANを被写体ブレの補正に流用できることを確認することに留まっている。実際に被写体ブレした画像は今回研究で用いた画像よりも多種多様なブレが存在する。そのため学習を行う際のカーネルを大きくすることでより大きいブレへの適応や、多様なブレに対応するためより多くのデータセットを用意する必要もあり、またそれらを実現するために学習に必要な計算資源を減らすためのネットワークの改良が必要になる。

参考文献

- [1] Ian J. Goodfellow et al, Generative Adversarial Nets.
arXiv:1406.2661v1, (Jun 2014).
- [2] Orset kupyn et al, DeblurGAN-v2:Deblurring(Orders-of-Magnitude)Faster
and Better. arXiv:1908.03826v1, (Aug2019)
- [3] S. Maeda, Unpaired Image Super-Resolution using Pseudo-Supervision
Conference on Computer Vision and Pattern Recongition. CVPR, pp.291-
300, (Feb 2020).
- [4] Xintao Wang et al, ESRGAN: Enhanced Super-Resolution Generative
Adversarial Networks. arXiv:1809.00219v2, (Sep 2018)
- [5] Kuan-yu Huang, esrgan-tf2: <https://github.com/peteryuX/esrgan-tf2>

謝辞

本研究の遂行および本論文の作成にあたり、懇切丁寧な御指導を賜りました出口利憲先生に心より感謝の意を申し上げます。