

## 修士論文要旨(2021 年度)

### 深層学習および超解像を用いた任意焦点画像の生成

### Digital Refocusing Using Deep Learning and Super-Resolution

電気電子情報通信工学専攻 相田 息吹

#### 1. まえがき

現在、デジタルカメラやスマートフォンのカメラといった通常のカメラによる撮影では撮影後に写真のフォーカス位置を変更することは不可能である。一方、多視点画像などの視点位置による視差を持つ複数枚の画像群を取得することで撮影後に任意焦点画像を合成するリフォーカス[1]という技術がある。しかし、これを実現するにはライトフィールドカメラや多眼カメラなどの特殊なカメラを用いた撮影が必要となる。これらのデバイスは構造の特殊さゆえに撮影画像の空間解像度を上げるのが困難であるという制約がある。一方で、視点位置による視差とは異なる三次元情報を持つ画像群としてフォーカスタックがある。フォーカスタックとは同じ被写体を異なる奥行きにフォーカスして撮影した画像群である。フォーカスタックは通常のカメラでも取得可能で数枚であればその撮影も容易であり、現在画像解析などを用いたフォーカスタックの研究が行われている。

そこで本研究では数枚の疎なフォーカスタックから深層学習を用いて密なフォーカスタックを生成する手法を提案する。本研究では畳み込みニューラルネットワーク(CNN)[2]、敵対的生成ネットワーク(GAN)[3]を使用してフォーカスタックの生成を行う。CNNは画像認識、画像生成分野で多くの功績を挙げておりフォーカスタックの生成にも応用できると考えられる。GANはニューラルネットワークを応用した生成アルゴリズムとして注目を集めており本研究ではフォーカスタックの生成に向け画像焦点を対象としたドメイン変換を行った。

#### 2. CNNを用いたフォーカスタック生成

本研究では名古屋大学の立岩史孝氏らのCNNモデル[4]を参考にフォーカスタック生成モデルを構築した。補完部分ではCNNのチャンネル数をフォーカスタックの枚数に見立て、畳み込み処理を繰り返しながらチャンネル数を増やすことで41枚の密なフォーカスタックを生成した。またネットワークの後半に超解像処理を取り入れることにより画像の高精度化を測った。図1にCNNモデルの全体構成を示す

##### 2.1 前処理

CNNモデルの学習に用いるフォーカスタックは現在公開されているものと数が少なく、密なものを実際に撮影することも難しい。そこでライトフィールド画像のデータセットからMATLABツールボックスを用いてフォーカスタックデータセットを作成した。また、作成できる数では学習データとして不十分であったため画像を $56 \times 56$ の大きさの任意領域に取り出し10000セットにした。また、入力をRGB分割することで30000セットのフォーカスタックを学習データとした。

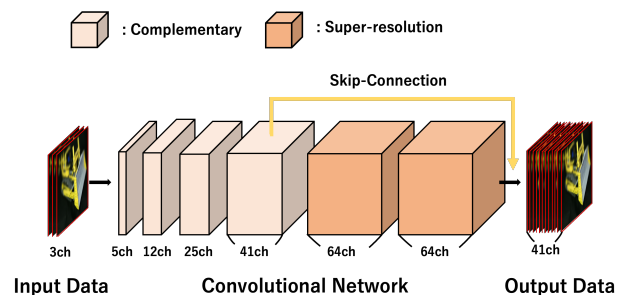


図1: CNNモデルの全体構成

## 2.2 超解像手法

本研究ではディープラーニングによる超解像手法としてVDSR、EDSRを用いた。これらはモデルを深層化するために残差学習という学習方法を用いている。残差学習は各層の出力値を最適化する通常のニューラルネットワークと異なり出力と入力との差分を最適化し、前の層の出力を伝搬する。それにより学習の効率化、勾配消失のリスク低減を行うことができる。モデル時実装時にはスキップ接続を取り入れることで残差学習を実現する。VDSRでは畳み込み処理を全て終えた後にスキップ接続を行い、EDSRでは各畳み込み層でスキップ接続を行う。

## 2.3 3枚入力でのフォーカスタック生成

本研究では遠近中焦点からなる3枚の疎なフォーカスタックから41枚の密なフォーカスタックを生成した。その際、フォーカスタックが焦点変更されているのかを主観評価で確認し、画像の精度をPSNR、SSIMによって客観評価をした。また、入力に用いる3枚のフォーカスタックの選び方についても検討を行った。図2、図3に生成画像の客観評価の結果を示す。

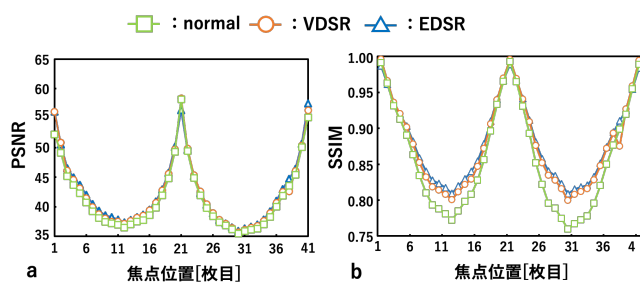


図2: フォーカスタック画像の評価結果(1, 21, 41 枚目)

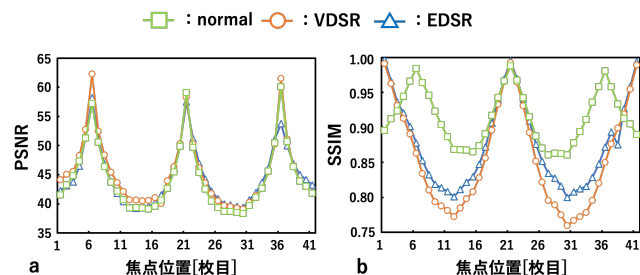


図3: フォーカスタック画像の評価結果(6, 21, 41 枚目)

図2では入力画像にフォーカスタックの1、21、41枚目を用いて、図3では6、21、41枚目を用いた。図2、3より入力部分から離れるほど画像の劣化が見られた。また、学習に用いる入力画像を中心に近づけ6、21、36枚目にする事で画像精度を向上させることができ、近づける前の超解像像を用いてないものと比較してPSNRの最小値を3.697、平均値を3.843上昇させることに成功した。SSIMにおいては最小値を10.85%、平均値を6.04%上昇させることに成功した。また、EDSR超解像を行った際の層の数の最適化も行った。PSNR、SSIMを用いて検証を行った結果スキップ接続を組み込む畳み込み層の数を10層にした際に最も高精度な画像を生成することができた。

## 2.4 2枚入力でのフォーカスタック生成

2枚の疎なフォーカスタックから41枚の密なフォーカスタックの生成も行った。評価は3枚入力時と同じ方法で行った。図4に生成画像の客観評価の結果を示す。主観評価では3枚入力と大きな差はないが図5より客観評価では画像精度の低下が確認された。これは学習の際に入力データセットから得られる画像の情報が少なくなっているためだと考えられる。

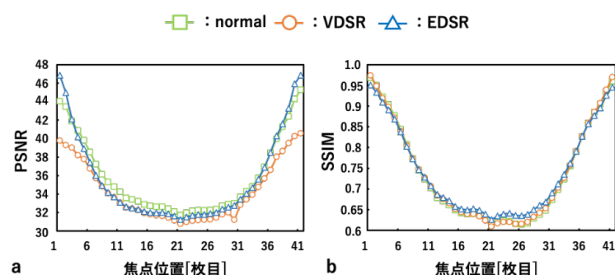


図4: 2枚入力の際のフォーカスタック評価結果

### 3. CycleGAN を用いた任意焦点画像生成

本研究では2枚の疎なフォーカスタックから41枚の密なフォーカスタックの生成に向けGANの一種であるCycleGANを用いた写真の焦点変更を行った。GANは偽物画像を生成するGeneratorと入力された画像が本物か偽物か判別するDiscriminatorの2つのニューラルネットワークから構成されている。GeneratorはDiscriminatorを欺けるように、Discriminatorは偽物画像と本物画像を見破れるようにそれぞれ交互に学習を繰り返し本物に近い画像の作り方を学んだ生成モデルを作る。

#### 3.1 CycleGAN モデル

CycleGANはGANの一種で画像のドメイン変換に特化している。図5にCycleGANの概要を示す。CycleGANではGeneratorに画像を入力しそれを異なるドメインの画像に変換する。

また、CycleGANではサイクル一貫性損失というものを導入し目的関数を定義することで、画像のドメイン変換を可能にする。Generatorの学習とDiscriminatorの学習を交互に繰り返し目的関数を満たすニューラルネットワークのパラメータを更新していく。以下にCycleGANの目的関数の式を示す。

$$G^*, F^* = \arg \min_{G, F} \max_{D_X, D_Y} \mathcal{L}(G, F, D_X, D_Y) \quad (1)$$

CycleGANではGeneratorとDiscriminatorが式(1)を満たすようにパラメータを更新していく。

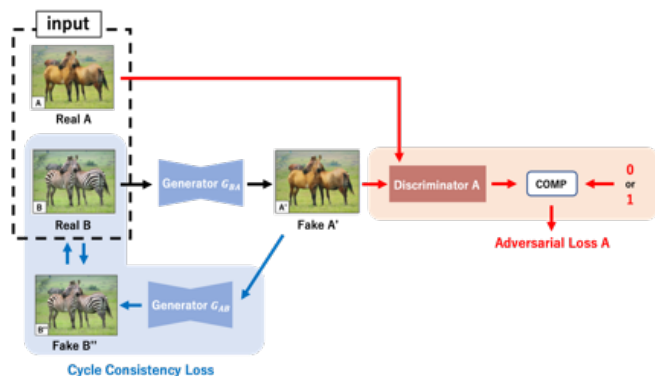


図5: CycleGANの概要

#### 3.2 CycleGAN を用いた画像生成

本研究ではCycleGANモデルを用いて近焦点画像、遠焦点画像のドメイン変換を行い肉眼による主観評価とPSNR、SSIMによる客観評価を行った。学習の際にはCNNモデルと同じくMATLABを用いて作成したフォーカスタックから遠焦点と近焦点のものをそれぞれドメインA、ドメインBとして入力した。図6にドメイン変換を行い生成した画像を示す。Real A、Real Bは遠焦点・近焦点の元画像を示し、Fake A、Fake Bは元画像から変換した際の遠焦点・近焦点画像を示す。街画像、ブルドーザ画像の両方において変換した際に画像の色味に変化が見られた。また、画質の劣化が強く焦点の変化を確認することができなかった。CycleGANモデルで学習を行なった際、焦点ではなく色味を特徴量として学習していると考えられる。また、表1にPSNR、SSIMの結果を示す。客観評価においても画質の劣化が大きいことが確認できた。そこで入力画像をRGB(Red, Green, Blue)の三種類に分割し、グレースケールで学習を行うことで色味を特徴量として学習してしまうことを避けられるのではないかと考えた。図7に入力画像をRGB分割して学習した際の生成画像を示す。

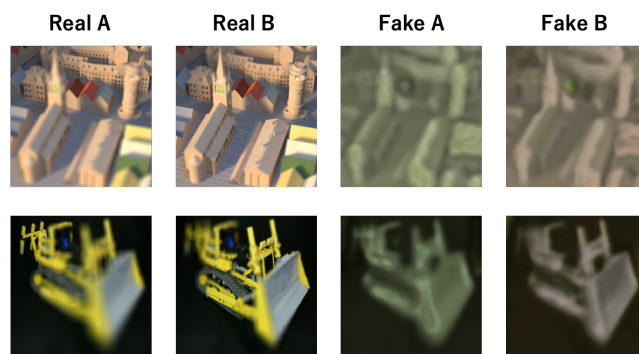


図6: CycleGAN生成画像

	Fake A	Fake B		Fake A	Fake B
Town	17.678	16.028	Town	0.6517	0.5993
Bulldozer	17.725	16.116	Bulldozer	0.6969	0.5711

a

b

表1: CycleGAN生成画像の客観評価

(a) PSNR (b) SSIM

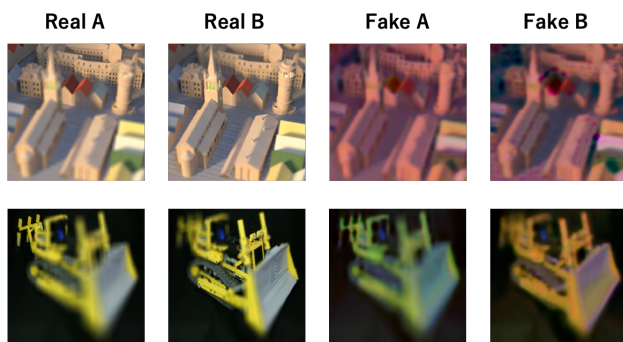


図 7:RGB 分割時の CycleGAN 生成画像

	Fake A	Fake B		Fake A	Fake B
Town	15.020	15.827	Town	0.6493	0.7833
Bulldozer	19.680	20.455	Bulldozer	0.7293	0.8688

a

b

表 2:RGB 分割時の CycleGAN 生成画像の客観評価

(a) PSNR (b) SSIM

RGB 分割をして学習を行った場合、街画像ではドメイン間の色変換が行われておらず焦点に変化が見られるため特徴量は正しく取れるようになっていると考えられる。しかし、ブルドーザーの画像では色味の変換も行われており、画像の劣化は両方で確認される。表 2 に客観評価の結果を示す。客観評価においても画像の劣化が確認された。

#### 4. おわりに

本研究では CNN および GAN を用いて疎なフォーカスタックから密なフォーカスタックを生成した。CNN モデルでは超解像手法、入力画像の最適化を行った。その結果 EDSR 超解像を用いてスキップ接続を行う層を 10 層にすること、入力に使うフォーカスタックを中心に近いものにすることで高精度の画像生成を行うことができるようになった。また、より少ない枚数からフォーカスタックを生成する手法について検証した。その結果、肉眼では元画像と変わらないものを作ることができたが PSNR、SSIM を用いた客観評価では画像精度の低下が確認された。

今後他の超解像手法やネットワーク構造を検討する必要があると考えられる。

CycleGAN モデルではカラー画像を入力として学習した場合、RGB 分割した画像を入力として学習した場合について焦点変更画像の生成を行った。生成画像を評価した結果、画像の劣化が確認された。これは 1 対 1 ドメインの変換という CycleGAN の構造上、学習に用いる画像が遠焦点、近焦点の 2 枚になり画像情報が少なくなるためだと考えられる。また、画像を分割したものを入力しているためドメインとして焦点の変化を認識することが難しいと考えられる。今後は焦点の変化がわかりやすい大量のデータセットを用意することや多値のドメイン変換が可能なモデルを検討する必要があると考えられる。

#### 参考文献

- [1] A. Isaksen et al., "Dynamically reparameterized light fields," Proceedings of the 27th annual conference on Computer graphics and interactive techniques, pp. 297– 306, ACM Press/Addison-Wesley Publishing Co., 2000 .
- [2] David E. Rumelhart; James L. McClelland, "Learning Internal Representations by Error Propagation," in Parallel Distributed Processing: Explorations in the Microstructure of Cognition: Foundations , MIT Press, 1987, pp.318-362.
- [3] 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)
- [4] 立岩 史孝, 伊藤 大貴, 高橋 桂太, 藤井 俊彰, "CNN を用いたフォーカスタック補間による任意焦点画像の合成," 3次元画像コンファレンス 2019 講演論文集 5-3, July. 2019