5M-08

セマンティックセグメンテーションネットワークを用いた超解像

廣瀬　理陽†　　藤田　悟†
法政大学　情報科学部†

1. はじめに

近年，画像の高解像度化を目指す超解像技術は，CNN(畳み込みニューラルネットワーク)を用いることでその性能が向上しており，多くの超解像 CNN が提案されてきた．これらはネットワークの構造が異なり，それにより特に高い性能を示す画像領域も異なる．そこで本研究では，画像領域ごとに複数の超解像 CNN の出力を重ね合わせる統合型超解像ネットワークを提案する．

2. 先行研究

2.1 超解像 CNN

CNN を用いた超解像は，CNN が入力画像の特徴を抽出し，それを基に高解像度化された画像を生成することによって実現している．これまで 3 層の畳み込み層で構成された SRCNN[1]，GAN による損失関数を用いた SRGAN[2]，畳み込み層によって抽出した特徴量に重みづけをする RCAN[3]など，特に画像の高周波成分の再現に注力したモデルが提案されている．モデルにより高い性能を示す画像領域は異なり，SRGAN は物体表面の質感，RCAN は輪郭の再現度が高い．

2.2 セマンティックセグメンテーション

画像内の各ピクセルに対しカテゴリ分類をするタスクにおいても CNN を用いた U-Net[4]が提案されている．U-Net は入力画像のサイズを縮小し，その後拡大を行う Encoder-Decoder 構造に加え，縮小前の特徴量を拡大後の特徴量に連結させることにより，画像内の物体の位置情報を保持した，より精密な領域でのカテゴリ分類ができる．

3. 提案手法

本章では，U-Net の出力に基づいて，ピクセルごとに超解像 CNN を使い分けるモデルを提案する．

3.1 モデルの全体像

今回提案するモデルの全体像を図 1 に示す．まず，低解像度の入力画像(LR)を，事前学習済みの 3 つの超解像 CNN に入力する．今回超解像 CNN として，2.1 節で取り上げた SRCNN，SRGAN，RCAN の 3 つを採用する．そして，入力画像の各ピクセルに対し，どの超解像 CNN の出力をより使用したほうが良いかを判断させるために，U-Net を用いて各ピクセルにおける 3 つの超解像 CNN の出力に対する重みを算出する．

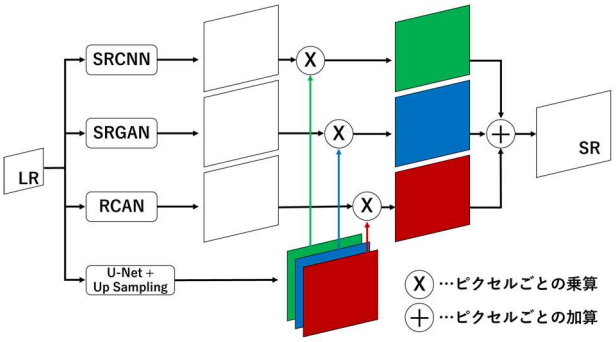

図 1. 提案モデルの全体像

U-Net は入力と同じサイズのデータを出力するが，超解像 CNN の出力サイズと合わせるために，U-Net の最後にアップサンプリング層を追加する．これらを合わせて「U-Net + Up Sampling」と呼ぶこととする．

この「U-Net + Up Sampling」の出力を，各超解像 CNN に対する空間方向のアテンションとし，それらを超解像 CNN の出力に，要素ごとにかける．そのようにして重みづけされた 3 枚の超解像画像をピクセルごとに足し合わせることで，最終的な出力画像(SR)を生成する．

3.2 U-Net + Up Sampling

「U-Net + Up Sampling」では 3 枚の特徴マップを出力させる．この 3 枚の特徴マップ出力をそれぞれ SRCNN，SRGAN，RCAN に対応させ，各ピクセル値をその超解像ネットワークの使用推奨度とみなす．その後，これら重みをそれぞれ 3 つの超解像 CNN の出力画像にかけることで，重みづけされた 3 枚の超解像画像を生成する．なお，提案モデルに組み込む U-Net において，画像の縮小を行う Max Pooling の層数を 2 層とした．

3.3 画像の統合

これまでの処理により生成した，3 枚の重みづ

Image Super-Resolution Using Semantic Segmentation Network
† Michiaki Hirose, Satoru Fujita, Faculty of Computer and Information Sciences, Hosei University

けされた超解像画像をピクセルごとに加算し，最終的な出力画像を得る．

これまでの操作は，以下のような数式として表すことができる．

$$I_{SR} = W_{SRCNN} \cdot SRCNN(I_{LR}) + W_{SRGAN} \cdot SRGAN(I_{LR}) + W_{RCAN} \cdot RCAN(I_{LR}) \quad (1)$$

I_{SR} は超解像出力画像，I_{LR}は低解像度入力画像を表す．また$SRCNN(\cdot)$，$SRGAN(\cdot)$，$RCAN(\cdot)$ はそれぞれ，SRCNN，SRGAN，RCAN を使用した超解像を表している．Wは U-Net で生成される，各超解像 CNN に対応した重みを意味している．

4. 実験

本章では，提案モデルと既存手法である SRCNN, SRGAN，RCAN に加え，ニューラルネットワークを使用しない手法として Bicubic 補間との性能比較を行う．画像の拡大倍率については 2，3，4 倍の 3 種類を実験した．

4.1 ネットワークの学習

各ネットワークの学習には DIV2K 学習用データセットを使用し，学習回数は，RCAN は 3000 エポック，それ以外のモデルは 1000 エポックとした．パラメータの最適化手法には Adam を使用し，学習率は，RCAN では 1000 エポックごとに10^{-4}，5×10^{-5}，10^{-5}と切り替え，RCAN 以外のモデルでは 800 エポックまで10^{-4}とし，残り 200 エポックは10^{-5}とした．

4.2 評価方法

出力画像の評価には，生成画像がどの程度原画像を再現できているかを表す PSNR(ピーク信号対雑音比)を用いる． PSNR は以下の式により算出する．

$$PSNR = 10 \cdot log_{10}\left(\frac{255^2}{MSE}\right) \quad (2)$$

MSE は 2 画像の輝度値の平均二乗誤差を表し，この誤差が小さければ，PSNR は大きな値となる．

4.3 実験結果

表 1 に，各データセットにおけるそれぞれの超解像モデルの生成画像と，その正解画像との PSNR の平均値をまとめる．評価データセットとして Set5，Set14，DIV2K 検証用画像，Urban100 を使用した．各倍率・データセットにおいて，1 位を太字，2 位を下線とともに示す．

Set5 に対しては全倍率において最も高い PSNR を記録した．4 倍拡大においては，Set14 以外のデータセットに対し，最も高い PSNR を記録した．

表 1. 検証用データセットにおける各超解像モデルの平均 PSNR.

倍率	モデル	データセット			
		Set5	Set14	DIV2K	Urban100
2倍	Bicubic	31.7716	28.2894	31.0333	25.4308
	SRCNN	33.9633	**29.8507**	**32.2911**	27.0555
	SRGAN	31.4065	27.5338	29.9095	26.6505
	RCAN	<u>34.5818</u>	27.8521	31.9059	**29.2371**
	提案モデル	**34.6918**	<u>28.3008</u>	<u>32.0653</u>	<u>29.0181</u>
3倍	Bicubic	28.6198	25.8275	28.2531	23.0284
	SRCNN	30.2451	26.7871	**29.0475**	24.0748
	SRGAN	28.4166	25.1323	26.8067	23.5825
	RCAN	<u>31.4079</u>	**26.9754**	28.9531	<u>25.6465</u>
	提案モデル	**31.5116**	<u>26.9478</u>	<u>29.0316</u>	**25.6569**
4倍	Bicubic	26.6522	24.2082	26.6780	21.6977
	SRCNN	28.0488	**25.1036**	<u>27.3138</u>	22.5176
	SRGAN	26.4511	23.7040	25.2118	21.9556
	RCAN	<u>28.9559</u>	24.1397	27.2374	<u>22.7054</u>
	提案モデル	**29.0409**	<u>24.2864</u>	**27.3482**	**22.9048**

5. まとめ

本研究では，セグメンテーションネットワークの出力に基づき，複数の超解像 CNN の出力を重ねる，統合型超解像ネットワークを提案した．実験の結果，特に 4 倍拡大超解像において，超解像 CNN を単体で使用する時と比べて高い性能を示すことができた．

参考文献

[1] Dong, C., Loy, C.C., He, K., Tang, X, ″Learning a deep convolutional network for image super-resolution.″, ECCV(2014).

[2] Christian Ledig, Lucas Theis, Ferenc Huszar, Jose Caballero, Andrew Cunningham, Alejandro Acosta, Andrew Aitken, Alykhan Tejani, Johannes Totz, Zehan Wang, Wenzhe Shi, “Photo-Realistic Single Image Super-Resolution Using a Generative Adversarial Network.”, CVPR(2017).

[3] Yulun Zhang, Kunpeng Li, Kai Li, Lichen Wang, Bineng Zhong, and Yun Fu, “Image Super-Resolution Using Very Deep Residual Channel Attention Networks.”, ECCV(2018).

[4] Olaf Ronneberger, Philipp Fischer, Thomas Brox, “U-Net: Convolutional Networks for Biomedical Image Segmentation.”, MICCAI(2015).