RI-005

4 出力の畳み込みニューラルネットワークを用いた超解像
Super-Resolution with Four Output Convolutional Neural Networks

加藤 裕† 大谷 真也† 黒木 修隆† 廣瀬 哲也† 沼 昌宏†
Yu Kato Shinya Ohtani Nobutaka Kuroki Tetsuya Hirose Masahiro Numa

1. はじめに

近年，ディジタルカメラ，携帯型情報端末，4K テレビなど，あらゆる映像機器の高解像度化が進んでいるが，高解像度の表示機器に低解像度の画像を入力する際には，何らかの解像度変換技術が必要となる．従来の線形フィルタによる補間法では，拡大後の画像にボケやジャギーといった画像の劣化が生じる問題がある．一方，元の画像に含まれない高周波成分を推定・復元する技術は，「超解像」と呼ばれ研究が盛んに行われている．Freeman ら [1] は補間法によりアップサンプリングされた低周波成分画像に対して高周波成分を加える手法を提案した．低周波成分に対応する高周波成分については，多くの事例をデータベース化し，辞書として保持するため，このような手法は事例参照型超解像と呼ばれる．以来，この枠組の手法は数多く提案されているが [1]–[8]，その中でも近年はスパースコーディングを取り入れた A+ [8] が高い性能を示している．

一方，Chao ら [9],[10] はニューラルネットワークを用いた超解像 (SRCNN) を提案し，驚異的に画質を向上させた．この手法では，Bicubic 法によって仮拡大した低周波成分画像から畳み込みニューラルネットワーク (CNN) を用いて，対応する高周波成分画像を生成する．事前学習においては，低周波成分画像を入力，高周波成分画像を教師信号とすることにより，end-to-end で CNN を訓練する．ただし，この手法には 1 つの問題点がある．CNN にはアップサンプリングの能力がないため Bicubic 法による仮拡大処理が必要となるが，Bicubic 法は CNN の学習プロセスに含まれていないため，そのフィルタ係数は最適化されない．この場合，初段の Bicubic 法は単なる次元拡大に過ぎず，高周波成分の生成には関わっていない．拡大後に CNN を動作させることは，メモリコストおよび計算コストの観点から非常に効率が悪い．

そこで，本研究では，Bicubic 法を用いずに CNN のみで超解像を実現するための新たなアーキテクチャを提案する．第 2 章で従来の SRCNN について説明したのち，第 3 章で CNN を用いた 3 つのアーキテクチャを提案する．第 4 章で様々な超解像手法との比較実験を行う．

2. 従来手法

本章では Chao ら [9],[10] が提案する従来の CNN を用いた超解像（SRCNN）の概要と問題点について述べる．

2.1. 概要

図 1 に SRCNN の概要を示す．入力画像を Bicubic 法で拡大したのち，3 層の CNN を通して高周波成分を推定する．Bicubic 法により拡大された低周波画像 $\boldsymbol{Y}$ に対し，1 層目の特徴マップ F_1 は，

$$F_1(\boldsymbol{Y}) = \max\left(0, W_1 * \boldsymbol{Y} + B_1\right) \tag{1}$$

のように求められる．ここで，W_1 は畳み込みのためのフィルタ群，B_1 はバイアスである．W_1 は 4 次元のテンソルであり，そのサイズは（入力チャンネル数 × フィルタの高さ × 幅 × 出力チャンネル数）となる．つまり，1 層目のフィルタサイズを $f_1 \times f_1$，特徴マップ数を n_1 とすると，W_1 のサイズは $1 \times f_1 \times f_1 \times n_1$ である．2 層目の特徴マップ F_2 は，

$$F_2(\boldsymbol{Y}) = \max\left(0, W_2 * F_1(\boldsymbol{Y}) + B_2\right) \tag{2}$$

のように求められる．W_2 のサイズは 2 層目のフィルタサイズを $f_2 \times f_2$，特徴マップ数を n_2 とすると，$n_1 \times f_2 \times f_2 \times n_2$ である．最後に，超解像画像は，

$$F(\boldsymbol{Y}) = W_3 * F_2(\boldsymbol{Y}) + B_3 \tag{3}$$

のように求められる．W_3 のサイズは $n_2 \times f_3 \times f_3 \times 1$ である．

学習段階では原画像群と，それらを Bicubic 法で縮小した後に再び拡大したボケ画像群を準備する．ボケ画像群を入力，原画像群を正解ラベルとして扱い，CNN の学習を行う．この学習により，CNN のフィルタ係数 W_1, W_2, W_3 およびバイアス B_1, B_2, B_3 を決定する．

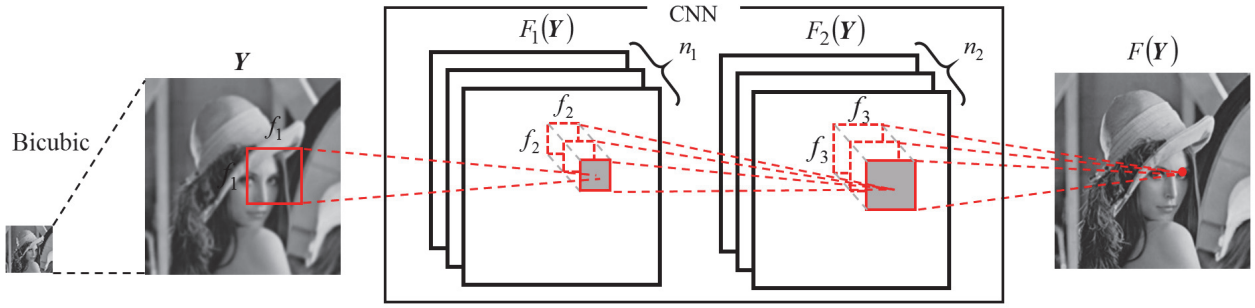


図 1 従来の SRCNN の流れ
Fig.1 Overview of conventional SRCNN.

†神戸大学大学院工学研究科,
Graduate School of Engineering, Kobe University

2.2. 問題点

従来の SRCNN では Bicubic 法を用いて仮拡大した画像に対して CNN を適用するが，この仕組みが 2 つの問題を生む．第一の問題は，Bicubic 法における 16 個の係数が，CNN に含まれていないため，最適化されないことである．図 2 に Bicubic 法における 2×2 倍拡大時の参照画素および出力画素の配置を示す. 出力画素数は入力画素数の 4 倍である．1 つの出力画素は 16 個の参照画素に対する重み付き線形和により求められる．ただし，A，B，C，D の位置によって 16 画素に対する重みが異なる．そのため，出力画像の奇数行，偶数行，奇数列，偶数列の画素は，異なる計算によって得られたものと言える. これによって図 3 に示すように，拡大画像に周期的な縞模様が発生する．後段の CNN はこれを補正する役割を負うことになる．

もう 1 つの問題は, 計算コストおよび利用メモリの増加である．例えば 2×2 倍拡大の場合，Bicubic 法で生成された画像は入力画像の 4 倍の大きさであるため，CNN も入力画像の 4 倍の大きさの特徴マップを全ての層で保持しなければならない．これはメモリを浪費するだけでなく，各層のフィルタサイズも相対的に大きくする必要があるため，計算コストの増加を引き起こす.

3. 提案手法

画像の 2×2 倍拡大を行うためには，1 つの入力画素あたり 4 つの画素を出力する構造が必要である．本章では，Bicubic 法を使用せずに CNN のみで超解像を実現するため，3 種類のアーキテクチャを提案する.

3.1. Type A: 4 並列の CNN (4P)

2 章で述べたように，出力画素は最も近い入力画素から見て (A) 左上，(B) 右上，(C) 左下，(D) 右下の画素群に分類することができる．我々は CNN がそれぞれの画素群に特化するべきと考え, すでに図 4 に示すような 4 並列の CNN を用いた超解像 (4P) を提案している [11].

この手法では，4 枚の画像がそれぞれ独立した CNN から生成される．例えば，CNN-A は画素群 (A) から構成される画像 $F_A(\boldsymbol{Y})$のみ生成する．画像 $F_A(\boldsymbol{Y})$は，

$$F_{A1}(\boldsymbol{Y}) = \max\left(0, W_{A1} * \boldsymbol{Y} + B_{A1}\right) \tag{4}$$

$$F_{A2}(\boldsymbol{Y}) = \max\left(0, W_{A2} * F_{A1}(\boldsymbol{Y}) + B_{A2}\right) \tag{5}$$

$$F_{A}(\boldsymbol{Y}) = W_{A3} * F_{A2}(\boldsymbol{Y}) + B_{A3} \tag{6}$$

と表される．なお，添字の A はその変数または関数が画素群 (A) 専用であることを示す．このとき，入力画像と

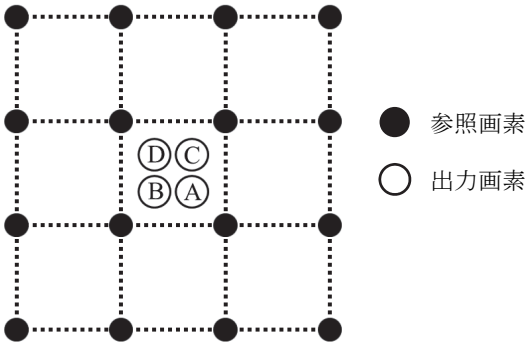


図 2　Bicubic 法における参照画素
Fig.2　16 Reference pixels of Bicubic method

図 3　Bicubic 法の係数に起因する縞模様
Fig.3　Periodical degradation derived from Bicubic filter

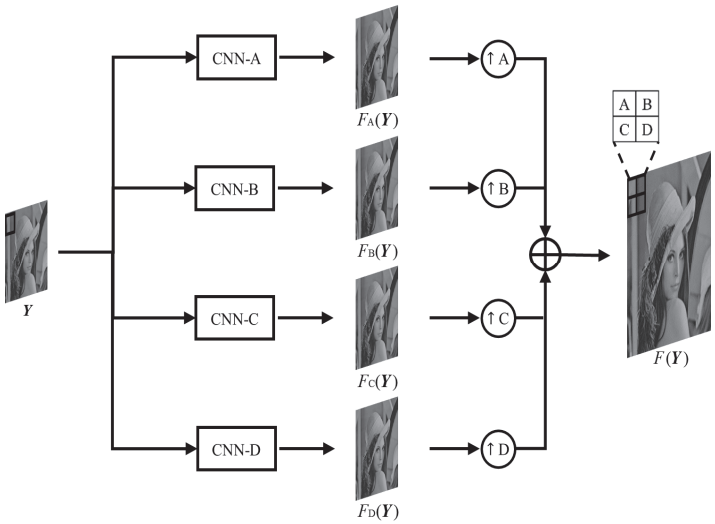


図 4　Type A: 4 並列の CNN
Fig.4　Type A: Four parallel CNNs

CNN-A の出力画像のサイズは同じである．同様に画素群 (B), (C), (D) に対応する CNN-B，CNN-C，CNN-D を独立に構成し，画像 $F_B(\boldsymbol{Y})$，$F_C(\boldsymbol{Y})$，$F_D(\boldsymbol{Y})$ を求める．最後に図 5 のように 4 枚の画像を再構成し，1 枚の超解像画像 $F(\boldsymbol{Y})$ を

$$F(\boldsymbol{Y}) = \uparrow \mathrm{A}(F_{\mathrm{A}}(\boldsymbol{Y})) + \uparrow \mathrm{B}(F_{\mathrm{B}}(\boldsymbol{Y})) + \uparrow \mathrm{C}(F_{\mathrm{C}}(\boldsymbol{Y})) + \uparrow \mathrm{D}(F_{\mathrm{D}}(\boldsymbol{Y})) \quad (7)$$

のように求める．ここで，$\uparrow \mathrm{A}(\boldsymbol{Y})$，$\uparrow \mathrm{B}(\boldsymbol{Y})$，$\uparrow \mathrm{C}(\boldsymbol{Y})$，$\uparrow \mathrm{D}(\boldsymbol{Y})$ は A~D それぞれの位置に対するアップサンプリングを示す．この手法は，Bicubic 法による事前拡大を必要とせず，拡大前の画像と拡大後の画像との間で完全に end-to-end の学習が可能になる．

3.2. Type B: 4 つの出力チャンネルを持つ CNN (4CH)

次に，4 つの出力チャンネルを持つ CNN (4CH) を提案する．処理の流れを図 6 に示す．Type A では画素群 (A), (B), (C), (D) をそれぞれ独立した 4 つの CNN から生成していたが，Type B では，それらの画素群を 1 つの CNN の 4 つの出力チャンネルから生成する．1 層目と 2 層目の出力は Type A と同様に

$$F_1(\boldsymbol{Y}) = \max(0, W_1 * \boldsymbol{Y} + B_1) \quad (8)$$

$$F_2(\boldsymbol{Y}) = \max(0, W_2 * F_1(\boldsymbol{Y}) + B_2) \quad (9)$$

によって求めるが，3 層目の出力のみ

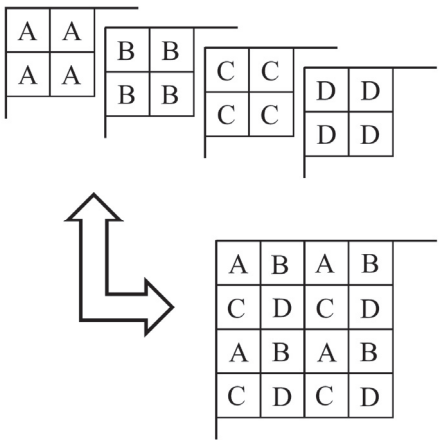


図 5　4 画素の分離と再構成
Fig.5　Separation and reconstruction of four pixels

$$F_3(\boldsymbol{Y}) = W_3 * F_2(\boldsymbol{Y}) + B_3 \quad (10)$$

のように求める．ここで，式 (10) の $F_3(\boldsymbol{Y})$ は，式 (3) や式 (6) の $F(\boldsymbol{Y})$ とは異なり，4 チャンネルである．つまり，W_3 のサイズは $n_2 \times f_3 \times f_3 \times 4$ である．このとき，$F_3(\boldsymbol{Y})$ の各チャンネルをそれぞれ $F_{3.1}(\boldsymbol{Y})$，$F_{3.2}(\boldsymbol{Y})$，$F_{3.3}(\boldsymbol{Y})$，$F_{3.4}(\boldsymbol{Y})$ とすると，超解像画像 $F(\boldsymbol{Y})$ は，

$$F(\boldsymbol{Y}) = \uparrow \mathrm{A}(F_{3.1}(\boldsymbol{Y})) + \uparrow \mathrm{B}(F_{3.2}(\boldsymbol{Y})) + \uparrow \mathrm{C}(F_{3.3}(\boldsymbol{Y})) + \uparrow \mathrm{D}(F_{3.4}(\boldsymbol{Y})) \quad (11)$$

と求められる．Type B は 1 つの CNN のみを用いるため，4 つの CNN を用いる Type A と比較して学習コストおよび処理時間を抑えることができる．

3.3. Type C: 回転平均を用いた 4CH (4CH-R)

次に前述の Type B に対して回転平均処理を導入する．実際の利用環境においては，たとえユーザが入力画像を 90 度，180 度，270 度に回転しても，同じ品質の超解像画像を出力することが望ましい．そこで，Type C では図 7 のように入力画像を 0°，90°，180°，270°に回転し，それぞれを Type B で拡大する．次に拡大後の画像を 0°，−90°，−180°，−270°に逆回転する．画像の回転を $R_\theta(\boldsymbol{Y})$ で表すと 4 枚の画像は，

$$F_\theta(\boldsymbol{Y}) = R_{-\theta}(F(R_\theta(\boldsymbol{Y}))) \quad (12)$$

と表される．この処理は CNN の内部構造を 0°, 90°, 180°, 270°回転させることと等価である．もし CNN のフィルタ係数が上下左右対称ならば，4 枚の画像は同一のものとなるが，バックプロパゲーションによる学習過程の中でフィルタが完全な対称形になることは極めて稀である．そこで最終的に 4 枚の画像を平均して，

$$\overline{F}(\boldsymbol{Y}) = \left(F(\boldsymbol{Y}) + F_{90^\circ}(\boldsymbol{Y}) + F_{270^\circ}(\boldsymbol{Y}) + F_{180^\circ}(\boldsymbol{Y})\right) / 4 \quad (13)$$

のように超解像画像を得る．Type C では入力画像の方向に依存しない安定した品質の画像を生成することが可能である．

4. 実験

本章では CNN のトレーニングを行った後，超解像処理後の画質について評価を行い，最後に最新の超解像技術を含む 7 手法について処理時間を含めた性能評価を行う．

4.1. 実験条件

比較手法は以下の 5 つである．

i)　Bicubic 法

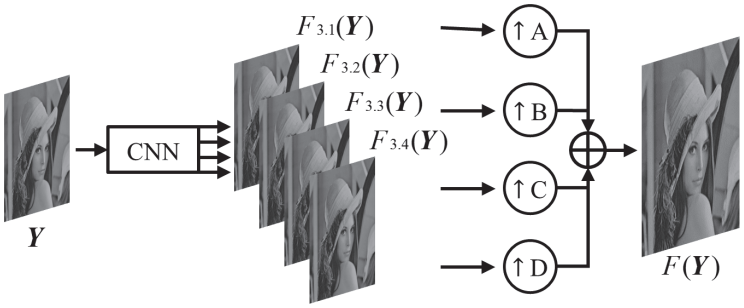


図 6　Type B: 4 つの出力チャンネルを持つ CNN
Fig.6　Type B: Single CNN with four output channels

ii)　従来手法：Bicubic 法 + CNN (SRCNN)
iii)　Type A：4 並列の CNN (4P)
iv)　Type B：4 つの出力チャンネルを持つ CNN (4CH)
v)　Type C：回転平均を用いた 4CH (4CH-R)

従来の SRCNN のフィルタサイズは文献 [10] に則り，$f_1=9$，$f_2=5$，$f_3=5$ とし，提案手法のフィルタサイズは，$f_1=5$，$f_2=5$，$f_3=3$ とした．いずれの手法も 1 層目と 2 層目の特徴マップ数は $n_1=64$，$n_2=32$ とした．

なお，超解像は画像の Y 成分 (輝度成分) にのみ適用し，評価も Y 成分について行う．画質評価には PSNR (Peak Signal to Noise Ratio) の値を用いる．全ての処理は CPU：Intel Core i7 3.50 GHz，メモリ：8.00 GB の PC 上で行う．

図 8　評価画像 (Set 5)
Fig.8　Image Set 5

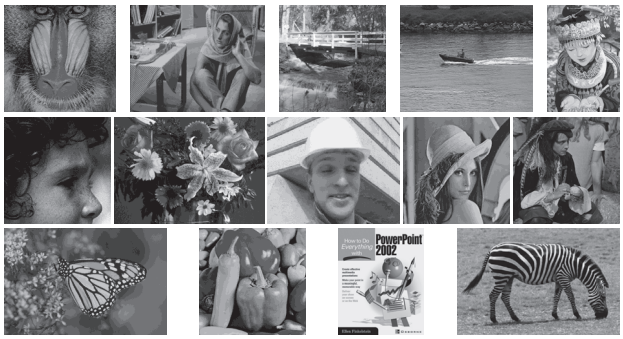


図 9　評価画像 (Set 14)
Fig.9　Image Set 14

4.2. CNN のトレーニング

まず 91 枚の画像 [10] を教師信号として CNN のトレーニングを行う．入力用の低解像度画像はこれら 91 枚を Bicubic 法により 1/2 倍に縮小したものとする．バックプロパゲーションは最大 1.6 億回とする．学習係数は 1 層目と 2 層目が 0.0001，3 層目が 0.00001 とする．

バックプロパゲーションに対する PSNR の変化を図 10 に示す．ここに示した PSNR は図 8 に示す評価画像 (Set 5) の PSNR の平均値である．33.66 dB の水平線は Bicubic 法の PSNR である．図 10 より，従来の SRCNN および 4P に比べ，4CH および 4CH-R は PSNR が速やかに上昇していることが分かる．4CH と 4CH-R ではフィルタサイズ f_1，f_2，f_3 が小さいことと，単一の CNN しか持たないことから，パラメータが速やかに収束したと考えられる．

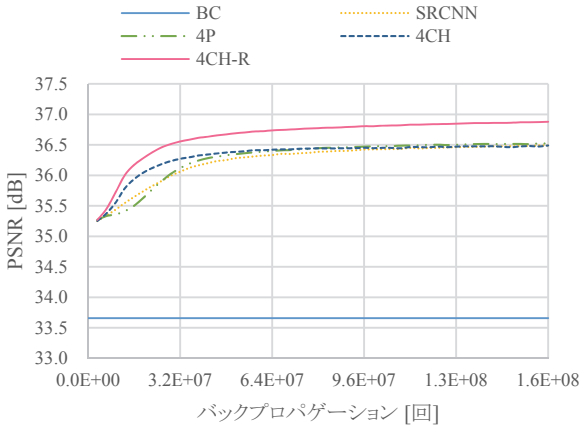


図 10　バックプロパゲーションの回数に対する PSNR の変化
Fig 10　Average PSNR curve for the number of back propagation

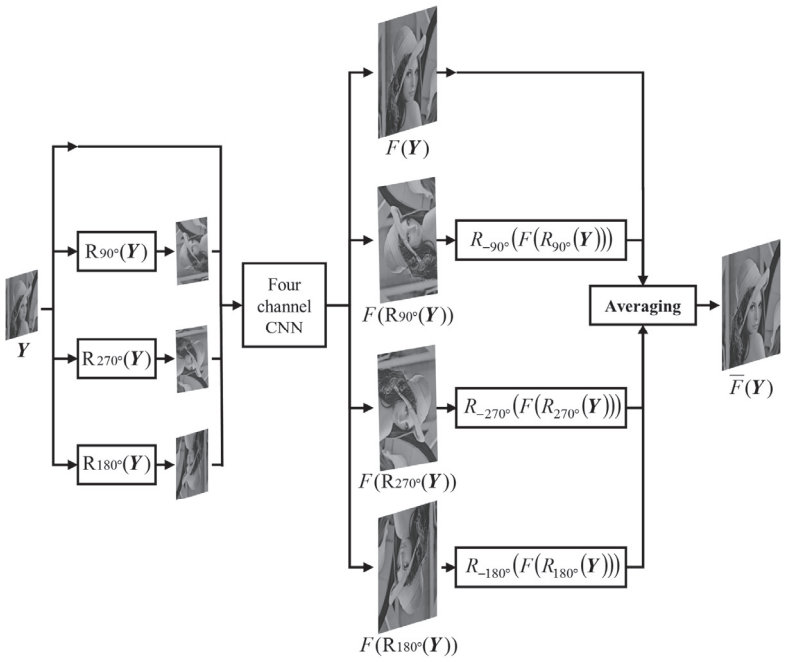


図 7　Type C: 回転平均を用いた 4 チャンネル出力の CNN
Fig.7　Type C: Four channel CNN with rotary averaging technique

表 1　PSNR 測定結果 (Set 5)

評価画像	PSNR [dB]				
	従来手法		提案手法		
	bibubic 法	SRCNN [10]	Type A : 4P [11]	Type B : 4CH	Type C : 4CH-R
baby	36.95	38.22	38.20	38.16	**38.38**
bird	36.81	40.80	40.75	40.72	**41.40**
butterfly	27.49	32.60	32.80	32.81	**33.29**
head	34.74	35.55	35.54	35.52	**35.62**
woman	32.30	35.27	35.32	35.25	**35.70**
平均値	33.66	36.49	36.52	36.49	**36.88**

表 2　PSNR 測定結果 (Set 14)

評価画像	PSNR [dB]				
	従来手法		提案手法		
	bibubic 法	SRCNN [10]	Type A : 4P [11]	Type B : 4CH	Type C : 4CH-R
baboon	24.86	25.65	25.66	25.66	**25.76**
barbara	27.91	28.42	28.39	28.32	**28.47**
bridge	26.63	27.86	27.85	27.84	**27.98**
coastguard	29.22	30.54	30.52	30.50	**30.61**
comic	25.80	28.14	28.16	28.19	**28.43**
face	34.71	35.53	35.51	35.50	**35.59**
flowers	30.16	32.88	32.95	32.93	**33.27**
foreman	35.43	38.95	38.92	38.90	**39.10**
lenna	34.55	36.37	36.39	36.37	**36.59**
man	29.14	30.76	30.78	30.78	**30.94**
monarch	32.73	37.12	37.27	37.26	**37.76**
pepper	35.00	36.96	36.94	36.95	**37.11**
ppt3	26.64	29.86	29.70	29.80	**30.17**
zebra	30.46	33.46	33.39	33.32	**33.76**
平均値	30.23	32.32	32.32	32.31	**32.54**

4.3. 超解像処理後の画質評価

評価画像は文献 [10] と同様に図 8 に示す 5 枚および図 9 に示す 14 枚とする．なお，これらの画像は学習用の 91 枚には含まれていない. これらを理想画像とし, Bicubic 法によって 1/2 倍に縮小した画像を超解像処理によって再び元のサイズに戻す.

図 11 に Set 5 の butterfly の処理結果の一部を示す. Bicubic 法による拡大画像 (b) では，全体がボケていることに加え, 斜めエッジの近辺に 2 画素周期の微かな縞模様が現れている．それに対して超解像による画像 (c)，(d)，(e)，(f) は鮮鋭であることが確認できる．以上の傾向は他の画像についても同様であった.

次に Set 5 および Set 14 の全画像の PSNR を，それぞれ表 1 および表 2 に示す．いずれの超解像手法も Bicubic 法に比べて 1dB〜3dB 程度高い PSNR を示すことが確認できる．従来の SRCNN，Type A : 4P，Type B : 4CH の 3 手法については，PSNR 上の大きな差は見られないが，言い換えれば，Type B : 4CH は省メモリのアーキテクチャでありながら, 他の 2 手法と同等の性能を維持していることが分かる．このことから，1 つの CNN の出力チャンネルを 4 つに分割する構成がメモリ削減の観点で高効率と言える.

一方，Type C : 4CH-R は，全画像について最高の PSNR を示した．これは回転平均の導入による効果が大きい．このことから，CNN 単体の特性は上下左右の方向に対して多少の不均衡があると考えられるが, それらを是正することで大きな画質改善につながることが分かる. 回転平均の導入は, 画像の入力方向に依存しない品質が得られる点でも大きな意義がある.

4.4. 各種超解像技術の性能比較

最後に，他の高性能な超解像技術 (ANR [7]，A+ [8]，SRCNN [10]) と提案手法 (4P，4CH，4CH-R) の性能を比較する．図 12 の横軸，縦軸はそれぞれ処理時間と Set 5 の PSNR の平均値である．図より，4CH が Bicubic 法を除いて最も高速であることが分かる．これは，4CH が 1 つの CNN のみで構成されており，さらにその内部のフィルタサイズが小さいためである．一方で 4CH-R が最も高い PSNR を示した．4CH-R の処理速度は 4CH に比べ 4 倍遅いが，SRCNN に比べ 1.2 倍速い結果となった．よって計算コストよりも品質を重視する場合は Type C : 4CH-R が有利である. なお, 4CH-R において画像の回転および平均に要する時間は極めて短く，CNN の処理に比べて無視できる程度であった. 以上の結果から我々が提案した 4 出力の CNN を用いたアーキテクチャの有用性が確認できた.

5. まとめ

本研究では，4 出力の CNN を用いた超解像アーキテクチャを提案した．従来の CNN を用いた超解像 (SRCNN) は Bicubic 法による仮拡大処理が必要であったが, 我々は,

Bicubic 法を必要とせず，直接 4 倍の画素を生成する超解像手法を実現した．実験では，Type B : 4CH が SRCNN よりも 4.3 倍速い処理速度を実現した．さらに，Type C : 4CH-R が従来 SRCNN に比べ PSNR で 0.39 dB 高い 36.88 dB を達成した．今後は 3×3 倍以上の拡大倍率に対応するとともに，カラー画像に対応する予定である．

参考文献

[1] W. T. Freeman, E. C. Pasztor, and O.T.Carmichael, “Learning low-level vision,” International Journal of Computer Vision, vol.40, no.1, pp.25-47, Oct. 2000.

[2] H. Chang, D. Y. Yeung, and Y. Xiong, “Super-resolution through neighbor embedding,” IEEE Conference on Computer Vision and Pattern Recognition, 2004.

[3] R. Timofte, V. D. Smet, and L. V. Gool, “Anchored neighborhood regression for fast example-based super-resolution,” IEEE International Conference on Computer Vision, pp.1920-1977, 2004.

[4] J. Yang, J. Wright, T. Huang, and Y. Ma, “Image super-resolution as sparse representation of raw image patches,” IEEE Conference on Computer Vision and Pattern Recognition, 2008.

[5] K. I. Kim and Y. Kwon, “Single-image super-resolution using sparse regression and natural image prior,” IEEE Transactions on Pattern Analysis and Machine Intelligence, vol.32, no.6, pp.1127–1133, 2010.

[6] J. Yang, J. Wright, T. Huang, and Y. Ma, “Image super-resolution via sparse representation,” IEEE Transactions on Image Processing, vol.19, no.11, pp.2861–2873, 2010.

[7] M. Bevilacqua, A. Roumy, C. Guilemot, and M. L. A. Morel, “Low-complexity single-image super-resolution based on nonnegative neighbor embedding,” British Machine Vision Conference, 2012.

[8] R. Timofte, V. D. Smet, and L. V. Gool, “A+: Adjusted anchored neighborhood regression for fast super-resolution,” Asian Conference on Computer Vision, 2014.

[9] C. Dong, C. C. Loy, K. He, and X. Tang, “Learning a deep convolutional network for image super-resolution,” European Conference on Computer Vision, 2014.

[10] C. Dong, C. C. Loy, K. He, and X. Tang, “Image super-resolution using deep convolutional networks,” Computer Vision and Pattern Recognition, 2014.

[11] 大谷 真也，加藤 裕，黒木 修隆，廣瀬 哲也，沼 昌宏，“4 並列の畳み込みニューラルネットワークを用いた超解像”，IEICE Transactions on Information and Systems，vol.J99-D，No.5，May 2016.

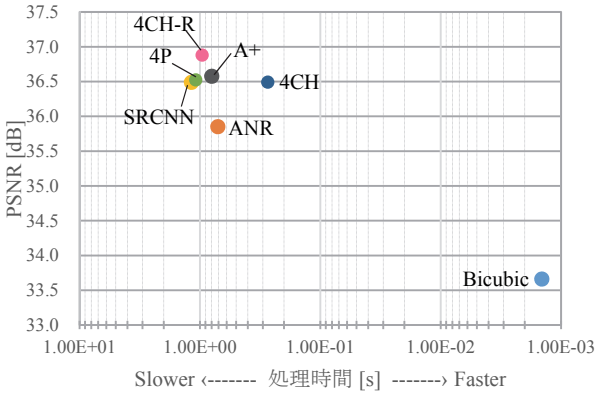
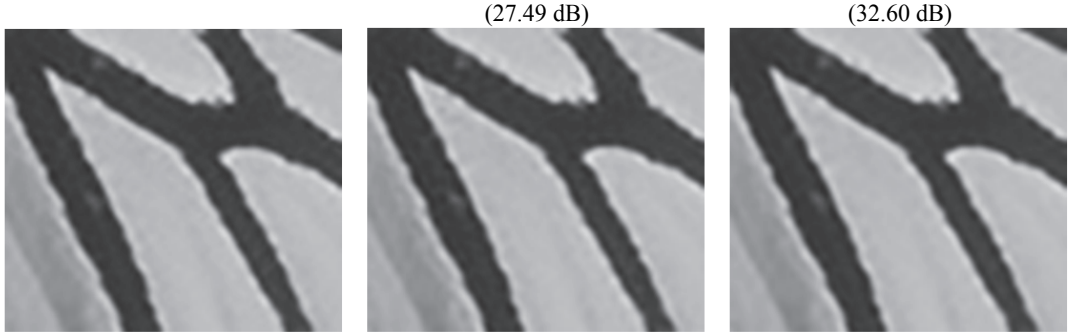


図 12　各種の超解像技術の性能比較

Fig.12　Performances of several super-resolution techniques

(a) 原画像　(b) Bicubic 法 (27.49 dB)　(c) 従来 SRCNN (32.60 dB)



(d) Type A: 4P (32.80 dB)　(e) Type B: 4CH (32.81 dB)　(f) Type C: 4CH-R (33.29 dB)

図 11　超解像処理後の“butterfly”の一部

Fig.11　A part of “butterfly” after super-resolution