情報処理学会論文誌 Vol.63 No.5 1216–1224 (May 2022)

各顔パーツを対象とした複数CNNモデルによる顔画像の高解像度化

丸井 勇輝[1,a)] 檜作 彰良[1,b)] 中山 良平[1]

受付日 2021年8月22日, 採録日 2022年2月4日

概要：CNN（Convolutional Neural Network）を用いた超解像技術は，顔画像の高解像度化にも応用されているが，特徴が大きく異なる各顔パーツを 1 つの超解像 CNN モデルで効果的に学習することは困難である．そこで本研究では，顔パーツごとに超解像 CNN モデルを構築し，それらの結果を合成することにより，顔画像を高解像度化する手法を開発した．実験試料は，CelebA Mask-HQ データセットに含まれる顔画像 30,000 枚と 8 つの顔パーツにアノテーションされたラベル画像である．本研究では，顔画像を高解像画像，1/4 に縮小した顔画像を低解像画像と定義し，提案手法により低解像画像から高解像画像を再構成した．各顔パーツの超解像 CNN モデルとして SRResNet を採用し，学習時の損失関数は，各 SRResNet の生成画像と高解像画像間の対象顔パーツ領域における平均二乗誤差で定義した．学習した各 SRResNet により低解像画像から各顔パーツの高解像画像を推定し，それらを合成することで顔画像の高解像度化画像（超解像画像）を生成した．そして，顔パーツごとに SRResNet を適用した提案手法と顔画像全体に SRResNet を適用した従来手法により生成した超解像画像に対し，主観的な知覚品質を評価するための観察者実験を行った．観察者実験では，提案手法と従来手法で生成された超解像画像 100 ペアを観察者に表示し，3 名の観察者が独立して，より高画質な画像を選択した．その結果，100 ペア中平均 92.33 ペアにおいて，提案手法の超解像画像が従来手法より高画質と評価され，提案手法の有用性が示唆された．

キーワード：超解像，顔画像，複数畳み込みニューラルネットワーク

Super Resolution Method Using Multiple Convolutional Neural Networks for Each of Facial Parts in Face Images

Yuki Marui[1,a)] Akiyoshi Hizukuri[1,b)] Ryohei Nakayama[1]

Received: August 22, 2021, Accepted: February 4, 2022

Abstract: Although super-resolution technique using CNN (Convolutional Neural Network) has been applied to improve the resolution of face images, it can be difficult for one super-resolution CNN model to effectively learn each face part with significantly different characteristics. The purpose of this study was to develop a super-resolution method using multiple CNNs for each of facial parts in face images. Our database consisted of 30,000 face images and the label images for eight facial parts in CelebA Mask-HQ dataset. In this study, a face image was defined as high-resolution image, whereas the down-sampled face image with a size of 1/4 was defined as low-resolution image. The high-resolution image was reconstructed from the low-resolution image with the proposed method. SRResNet was used for the super-resolution CNN model of each face part. The loss function was defined by the mean squared error in the target face part regions between the image generated with each SRResNet and the high-resolution image. In the proposed method, the high-resolution images of each face part were estimated from the low-resolution image by each learned SRResNet, and the super-resolution image of the face image was generated by synthesizing them. An observer study was conducted to evaluate the subjective perceptual quality of the super-resolution images generated by the proposed method applying SRResNet to each face part and the conventional method applying SRResNet to the entire face image. In the observer study, a pair of super-resolution images generated by both methods were displayed on a display monitor. Three observers independently selected one image considered as higher image quality from the pairs. In an average of 92.33 pairs out of 100 pairs, the super-resolution images generated by the proposed method were evaluated as higher image quality than the conventional method.

Keywords: super-resolution, face images, multiple convolutional neural networks

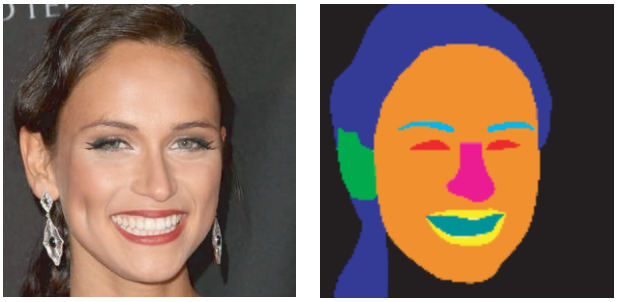
1. はじめに

画像，映像の解像度を後処理で向上する技術として超解像技術があり，事例ベース法 [1] やスパースコーディング法 [2] など画像処理に基づく手法が提案されてきた．近年では，畳み込みニューラルネットワーク（CNN：Convolutional Neural Network）を用いた深層学習に基づく超解像 CNN が多数報告されている [3], [4], [5], [6]．Dong らは，CNN を画像の高解像度化に初めて応用し，SRCNN（Super-Resolution CNN）を提案した [3]．SRCNN は 3 層の畳み込み層で構成され，1 層目は特徴抽出，2 層目は非線形マッピング，そして 3 層目は画像復元の役割を担う．SRCNN は，事例ベース法やスパースコーディング法と比較して，高品質な高解像度化画像を生成できることが報告されている．Kim らは，20 層の畳み込み層を積層した VDSR（Very Deep Super-Resolution）を提案した [5]．VDSR は，高解像画像ではなく，高解像画像と低解像画像の差である高周波成分を予測することで，効果的な学習を行った．また，Ledig らは，16 個の残差ブロックと pixel shuffler で構成される SRResNet を提案した [6]．SRResNet は，pixel shuffler で，特徴マップのチャネルを空間方向に並べ替えることにより，高解像特徴マップを出力し，より高品質な高解像度化画像の生成を可能にした．

顔画像を対象とした超解像 CNN モデルでは，顔画像特有の特徴に重みを置いた学習を期待する手法が報告されている [7], [8], [9], [10], [11]．Zhou らは，低解像顔画像と抽出した特徴マップを連結し，高解像度化画像を推定する Bi-channel CNN を開発した [7]．Chen らは，セマンティックセグメンテーションネットワークにより推定された各顔パーツのマスク画像と抽出した特徴マップを連結する FSRNet（Face Super-Resolution Network）[8]，Kalarot らは，推定された各顔パーツのアテンションマップと特徴マップを連結する CAGFace（Component Attention Guided Face Super-Resolution Network）を提案している [9]．また，Zhao らは，顔パーツのマスク画像を推定するネットワークを超解像 CNN に組み込んだ SAAN（Semantic Attention Adaption Network）を報告した [10]．これらの研究では，超解像 CNN モデルが各顔パーツに重みを与えた学習により，顔画像の高解像度化の精度を向上できることを示した．

一方，Ma らは各顔パーツのアテンションマップを連結した 1 つのネットワークで，各顔パーツの特徴を効果的に学習するには限界があると報告している [11]．また，これらの超解像 CNN モデルは顔画像全体の損失値に基づき学習するため，画素数が多いクラス（顔パーツ）に対して高解像度化の精度が高くなるように学習が進み，画素数が少ない顔パーツに対しては十分な学習ができない可能性がある．そこで本研究では，各顔パーツを対象とした超解像 CNN モデルをそれぞれ用意したネットワークを構築する．そして，対象顔パーツ領域の損失値で学習した各超解像 CNN モデルの出力を合成することにより，顔画像を高解像度化する手法を開発する．

[1] 立命館大学大学院理工学研究科
Graduate School of Science and Engineering, Ritsumeikan University, Kusatsu, Shiga 525–8577, Japan
[a)] ri0074rr@ed.ritsumei.ac.jp
[b)] hizukuri@fc.ritsumei.ac.jp



(a) Face image (b) Label image

図 1 顔画像とラベル画像の例
Fig. 1 Example of face and label images.

2. 実験試料

実験試料は，CelebA Mask-HQ データセット [12] に含まれる顔画像 30,000 枚と 8 つの顔パーツ（髪，眉，目，鼻，唇，歯，耳，皮膚）に手動でアノテーションされたラベル画像である（**図 1** 参照）．本研究では，ネットワークの学習用に 24,000 枚，検証用に 4,500 枚，そしてテスト用に 1,500 枚使用した．顔画像サイズは 512 × 512 画素であり，濃度分解能は 24 bit（RGB color）である．

本研究では，顔画像（512 × 512 画素）を高解像画像，1/4，のサイズにダウンサンプリングした顔画像を低解像画像（128 × 128 画素）と定義した．また，線形濃度階調変換により画素値を 0.0 から 1.0 の範囲に正規化した低解像画像はネットワークの入力として，高解像画像は学習時の教師画像および評価時の正解画像として用いた．ラベル画像は，学習時に各顔パーツ領域の損失値を求めるため，また評価時に各顔パーツを対象とした超解像 CNN モデルの出力を合成するために使用した．

3. 方法

3.1 ネットワーク構造

本研究では，各顔パーツを高解像度化する CNN モデルとして，16 個の残差ブロック [13] から構成される SRResNet [6] を用いた．**図 2** に，SRResNet のネットワーク構成を示す．図中の k，n，s は，それぞれ畳み込み層のカーネルサイズ，カーネル枚数，カーネルのストライド幅を表す．また LR，SR は，入力低解像画像，CNN モデルにより出力された超解像画像である．残差ブロックは，畳み込み

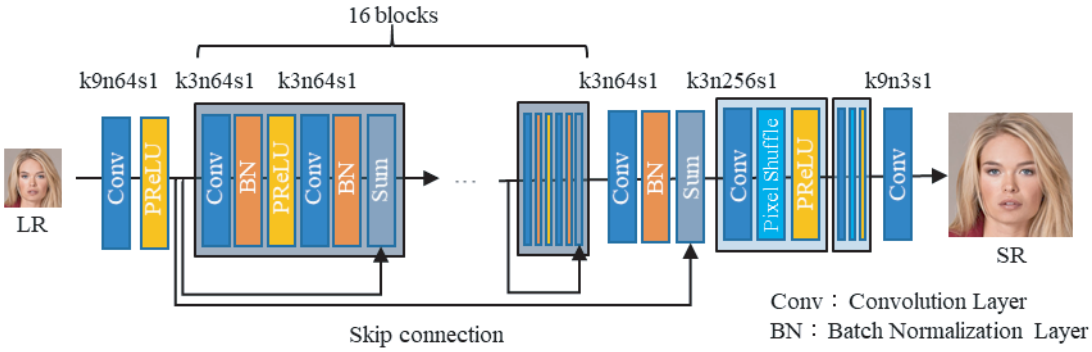


図 2 SRResNet のネットワーク構造
Fig. 2 Network architecture of SRResNet.

層，バッチ正規化層 [14] の 2 セットと，その間に PReLU（Parametric Rectified Linear Unit）活性化関数 [15] を追加した構成である．残差ブロックの各畳み込み層は，カーネルサイズ：3×3，カーネル枚数：64，カーネルのストライド幅：1 に設定し，入出力のデータサイズが変化しないようにパディングを行った．また，バッチ正規化層のモーメンタムは，0.99 に設定した．各残差ブロックの入力と出力は，勾配消失問題に対処可能な Skip connection [16], [17], [18] により連結される．また，SRResNet はネットワーク内でアップサンプリングを行っており，その手法として，特徴マップのチャネルを空間方向に並べ替える pixel shuffler [19] を導入している．

図 3 に，提案ネットワークの構成を示す．各 SRResNet により低解像画像から各顔パーツの高解像画像を推定し，それらを合成することで顔画像の超解像画像を生成する．

3.2 ネットワークの学習

顔パーツ c の超解像 CNN モデルの学習における損失関数 $loss_c$ は，生成画像 f_c と正解画像 y 間の対象顔パーツ領域における平均二乗誤差で定義した．

$$loss_c = \frac{1}{n_c} \sum_{l \in \text{each part region } c}^{n_c} (y_{cl} - f_l)^2 \qquad (1)$$

n_c は，顔パーツ c の領域内の画素数を表す．

SRResNet の学習パラメータは，ミニバッチサイズ：4，学習率：1×10^{-4}，モーメンタム：0.9，エポック数：10 とした．また，最適化手法として Adam を用いた．

3.3 評価方法

本研究では，評価用データの顔画像を 1/4 のサイズにダウンサンプリングした低解像画像をネットワークに入力し，超解像画像を生成した．超解像画像の正解画像に対する忠実度を評価する客観的な指標として，Peak Signal to Noise Ratio（PSNR）と Structural Similarity Index Measure（SSIM）を用いた．PSNR は，超解像画像が正解画像と比較して，どの程度劣化したかを評価する指標であり，PSNR が高いほど劣化が少ないことを示す．

$$\text{PSNR} = 20 * \log_{10}\left(\frac{MAX}{MSE}\right) \text{[dB]} \qquad (2)$$

ここで，MAX は正解画像の最大画素値を表す．また，MSE は，超解像画像と正解画像の平均二乗誤差である．一方，SSIM は，画素値，コントラスト，構造の 3 要素がどの程度変化したかを統合的に評価する指標である．

$$\text{SSIM} = \frac{(2\mu_x\mu_y + c_1)(2\sigma_{xy} + c_2)}{(\mu_x^2 + \mu_y^2 + c_1)(\sigma_x^2 + \sigma_y^2 + c_2)} \qquad (3)$$

μ_x，μ_y，σ_x，σ_y，σ_{xy} は，それぞれ正解画像 x と超解像画像 y の平均画素値，標準偏差および共分散である．また，c_1，c_2 は評価値が大きくなりすぎないよう制御するための定数であり，1×10^{-4} と 9×10^{-4} に設定した．

本研究では，提案手法の有用性を検証するため，顔画像全体に SRResNet を適用する高解像度化手法（SRResNet），また，FSRNet [8] を参考に，顔画像全体に適用する SRResNet の特徴マップに各顔パーツのマスク画像を連結した手法（SRResNet+Mask）と比較する．SRResNet，SRResNet+Mask の PSNR，SSIM に対する提案手法の有意差を評価するため，両側 t-検定を行った．ここでは，p 値が有意水準 5%未満を満たすとき，有意に差があると評価した．

画像の高解像度化において，客観的指標を用いた評価は人間の主観的評価としばしば異なる問題がある [20]．そこで，主観的な知覚品質を評価するため，顔画像の解析処理研究に従事する学生ボランティア 3 名が参加した Two-alternative forced choice 法 [21] による観察者実験を実施した．Two-alternative forced choice 法は，2 つの選択肢からどちらか 1 つを強制的に選択させる方法で，反応バイアス（観察者の経験に基づく正確性の違いによる誤差）を含みにくい利点がある．観察者実験では，顔パーツごとに SRResNet を適用した提案手法と顔画像全体に SRResNet を適用した従来手法により生成した超解像画像がモニタ上にペアで表示され，観察者は知覚品質を比較し，より高画質な超解像画像を選択した．提案手法と従来手法の超解像

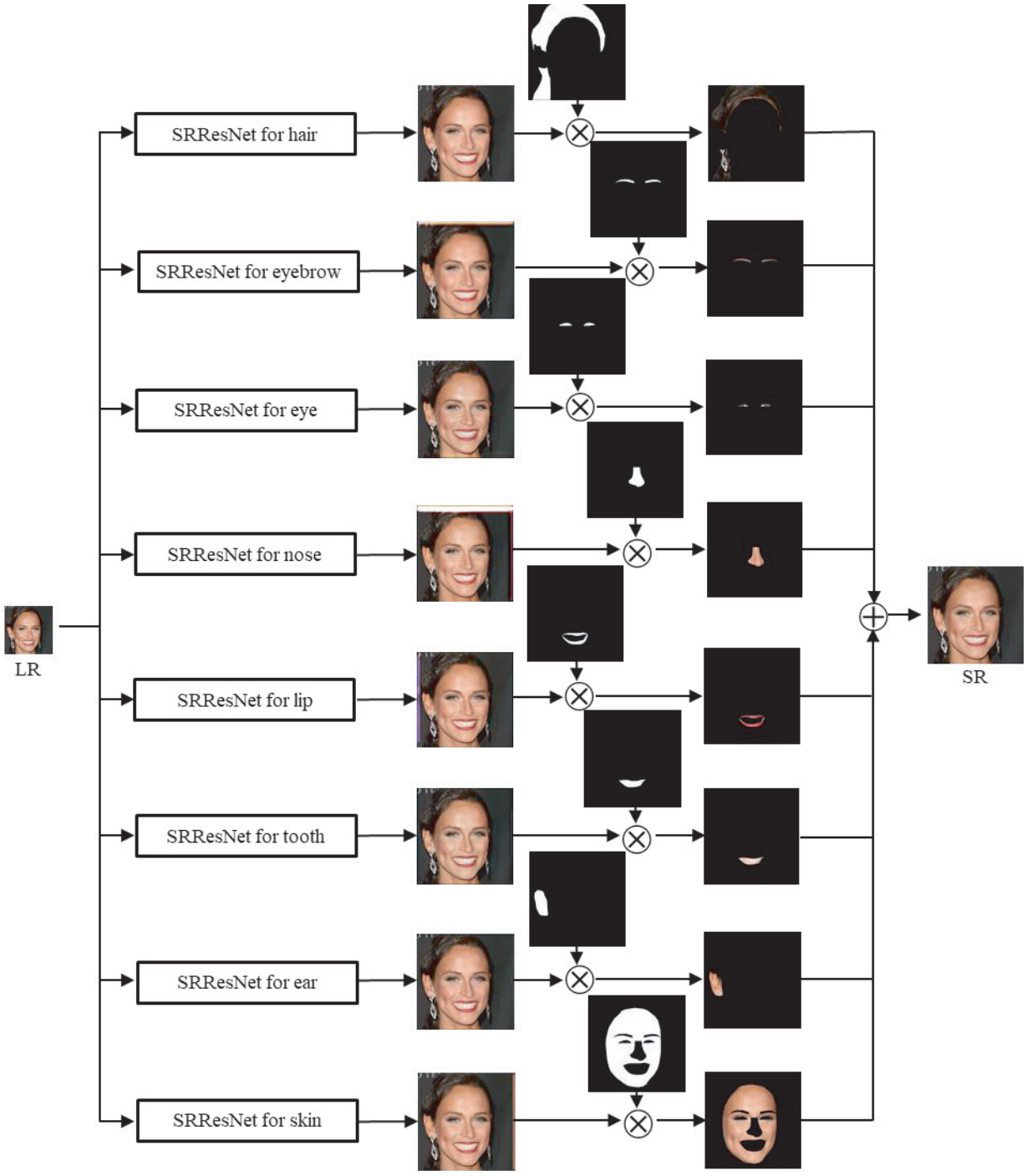


図 3 提案ネットワークの構成
Fig. 3 Overall scheme of the proposed network.

画像の表示位置は無作為に左右を入れ替え，表示画像を生成した手法に関する情報は観察者に提供しなかった．観察者実験では，100 ペアの超解像画像が比較され，高画質として選択された頻度を知覚品質スコアと定義した．

4. 結果

図 4 に，bicubic interpolation 法 [22] で拡大した低解像画像，高解像画像（正解画像）および SRResNet，SRResNet+Mask，提案手法で生成した超解像画像を示す．提案手法は，SRResNet や SRResNet+Mask と比べ，瞳孔や虹彩，歯の境界まで，より鮮明に再構成できたことから，その有用性が示された．

表 1 に，bicubic interpolation 法で拡大した低解像画像と 3 手法により生成された超解像画像の正解画像に対する忠実度を示す．すべての顔パーツにおいて，提案手法により PSNR が改善した．全顔パーツに対する提案手法の平均 PSNR は 46.89 dB で，SRResNet（46.29 dB，$p < 0.001$），SRResNet+Mask（46.65 dB，$p = 0.012$）より有意に高い

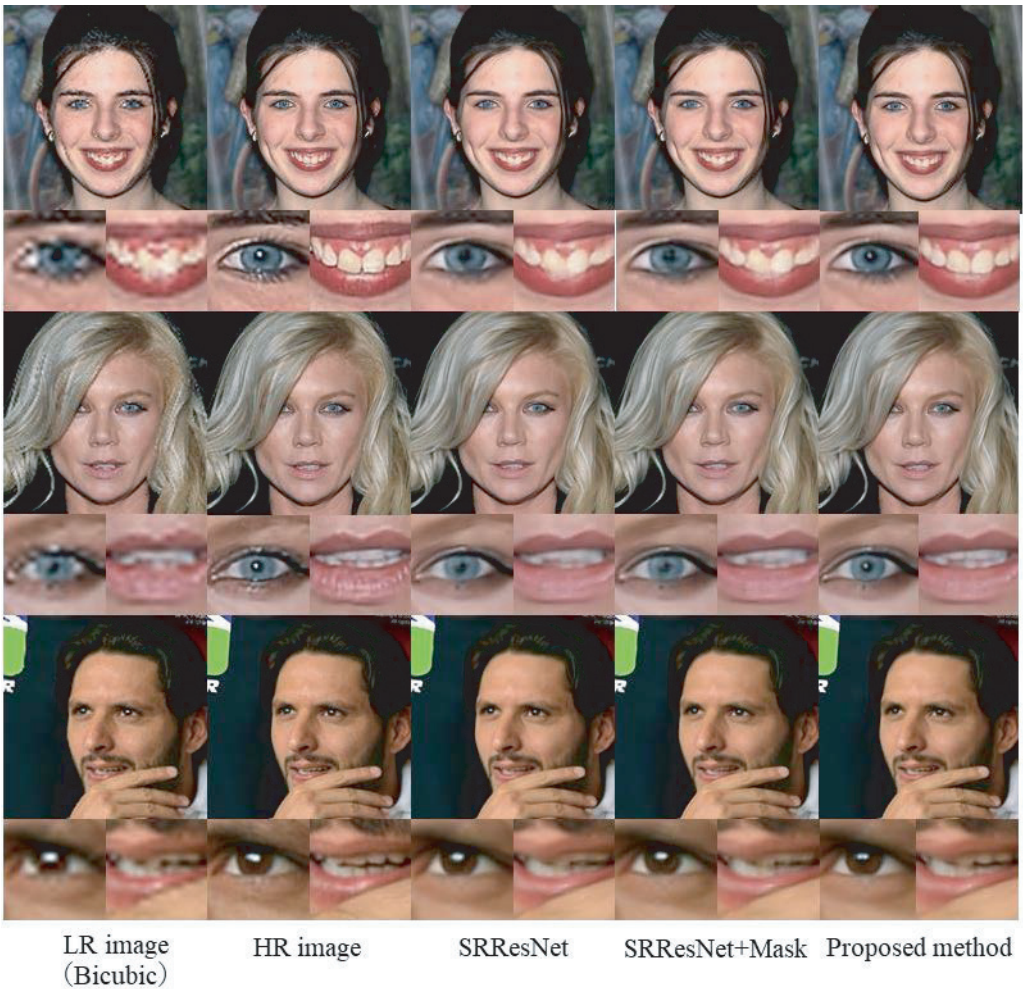


図 4 提案手法と従来手法による超解像画像の比較

Fig. 4 Comparison of the super-resolution images with the proposed method and the conventional methods.

表 1 提案手法と従来手法による超解像画像の忠実度の比較

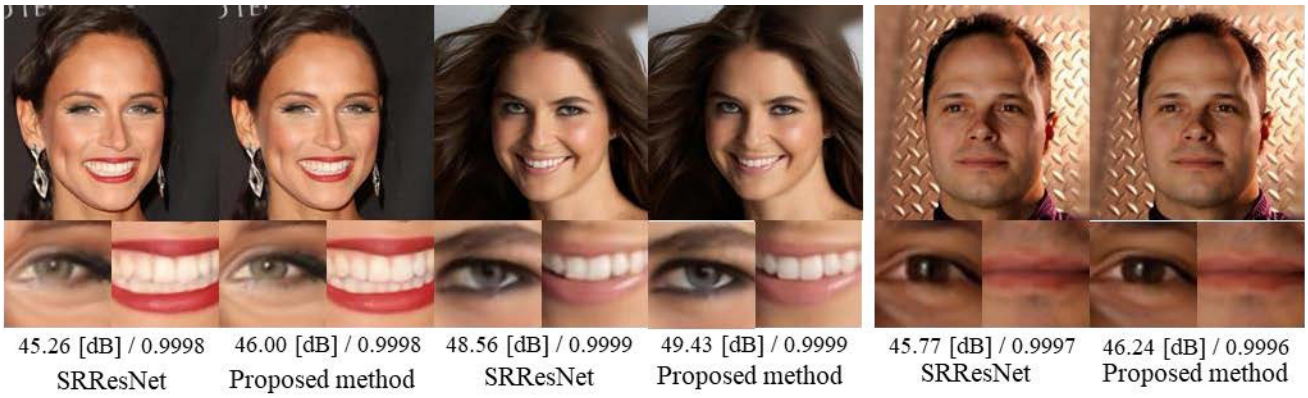
Table 1 Comparison of the fidelities of the super-resolution images with the proposed method and the conventional methods.

Parts	Bicubic	SRResNet	SRResNet + Mask	**Proposed method**
Hair	33.45±5.08 / 0.9996±0.0018	35.04±4.94 / 0.9993±0.0050	35.34±4.82 / 0.9996±0.0022	**35.36±4.77 / 0.9995±0.0017**
Eyebrow	49.90±4.42 / 0.9999±0.0010	51.00±4.21 / 0.9999±0.0011	51.19±4.19 / 0.9999±0.0008	**51.23±4.14 / 0.9999±0.0008**
Eye	45.02±3.67 / 0.9994±0.0041	46.76±3.56 / 0.9996±0.0035	46.90±3.50 / 0.9996±0.0033	**47.69±3.47 / 0.9996±0.0029**
Nose	50.74±3.95 / 0.9999±0.0003	51.87±3.13 / 0.9999±0.0003	52.83±3.55 / 0.9999±0.0002	**52.99±3.55 / 0.9999±0.0002**
Lip	48.32±3.39 / 0.9999±0.0004	50.04±3.13 / 0.9999±0.0004	50.32±3.15 / 0.9999±0.0004	**50.66±3.21 / 0.9999±0.0003**
Tooth	47.47±4.90 / 0.9998±0.0016	48.98±4.77 / 0.9998±0.0013	49.20±4.82 / 0.9998±0.0013	**49.60±4.77 / 0.9998±0.0012**
Ear	45.57±4.42 / 0.9999±0.0014	47.15±4.38 / 0.9999±0.0024	47.38±4.30 / 0.9999±0.0017	**47.41±4.13 / 0.9999±0.0010**
Skin	38.08±3.54 / 0.9999±0.0011	39.44±3.28 / 0.9999±0.0007	40.04±3.44 / 0.9999±0.0007	**40.22±3.48 / 0.9999±0.0008**
Ave.	44.82±4.17 / 0.9998±0.0014	46.29±3.92 / 0.9998±0.0018	46.65±3.97 / 0.9998±0.0013	**46.89±3.94 / 0.9998±0.0011**

PSNR: Ave.±Std.[dB] / SSIM: Ave.±Std.

結果となった．一方，提案手法の平均 SSIM は 0.9998 で，SRResNet（0.9998），SRResNet+Mask（0.9998）と同等であった．

提案手法と SRResNet で生成された超解像画像の知覚品質を評価する観察者実験の結果，各観察者の提案手法に対する知覚品質スコアは 100，97，80 で，SRResNet（0, 3, 20）に比べ，大幅に高かった．観察者 3 名の提案手法に対する平均知覚品質スコアは 92.33，SRResNet は 7.67 であ



(a) 提案手法による超解像画像が高画質と評価された例

(b) SRResNet による超解像画像が高画質と評価された例

図 5 観察者実験で高画質と評価された超解像画像の例

Fig. 5 Examples of super-resolution images considered as higher image quality in the observer study.

り，提案手法の有用性が示された．

5. 考察

顔画像全体に適用する SRResNet の特徴マップに各顔パーツのマスク画像を連結した SRResNet+Mask の PSNR は，SRResNet より高かったことから，顔画像特有の特徴に重みを置いた学習ができたと考える．しかし，SRResNet+Mask は，1 つの CNN モデルで構成されており，特徴が大きく異なる各顔パーツを効果的に学習できなかった可能性がある．また，顔画像全体の損失値に基づき学習するため，画素数が少ない小さな顔パーツに対して十分な学習ができなかった可能性もある．一方，提案手法では，顔パーツごとに超解像 CNN モデルを構築したことから，各 CNN モデルは類似した信号パターンを効率的に学習できたと考える．また，各 CNN モデルは対象顔パーツ領域の損失値に基づき学習したことから，小さな顔パーツに対しても十分な学習が可能となり，瞳孔や虹彩，歯の境界まで，より鮮明に再構成できたと考えられる．

図 5 に，観察者実験で観察者 3 名が高画質と評価した超解像画像の例を示す．提案手法による超解像画像は，従来手法に比べ，特に瞳孔や虹彩，歯の境界が鮮明であった．したがって，それらが写った画像に対して，提案手法の方がより高画質と選択される傾向にあった．一方，それらが確認できない虹彩が黒い画像や口を閉じた画像においては，知覚品質の判断が困難となり，従来手法による超解像画像が選ばれるケースもあった．したがって，目や歯の領域が，知覚品質に大きな影響を及ぼしたと考える．

CNN を用いた超解像モデルの損失関数として平均二乗誤差を用いた場合，高解像度化画像がボケる傾向にあることが報告されている [20]．提案手法は従来手法と同様に，損失関数として平均二乗誤差を採用した．しかし，提案手法による高解像度化画像は，従来手法と比較して，ボケが少ないことが確認できた．したがって，各超解像 CNN モデルが学習する信号パターンを限定されたことで，平均二乗誤差を使用した際に生じるボケを低減できたと考える．今後，提案手法の損失関数を平均二乗誤差から Perceptual loss [23] などに変更することにより，よりシャープな超解像画像を生成できる可能性がある．

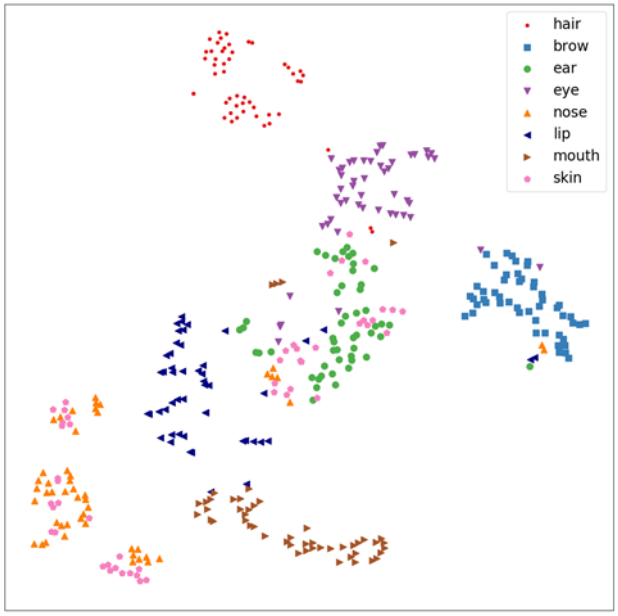


図 6 t-SNE による SRResNet の特徴マップの可視化

Fig. 6 Visualization of feature maps for SRResNet by t-SNE.

顔画像を学習した SRResNet の中間層の特徴マップに t-SNE [24] を適用することにより，各顔パーツの特徴マップを可視化した結果を**図 6** に示す．皮膚，鼻，耳の分布に重なりが見られるが，各顔パーツの特徴量がクラスタに分かれており，顔パーツ間に特徴の違いがあることが確認できた．したがって，本研究で 8 つの顔パーツに対し，SRResNet を構築したことは妥当であったと考える．

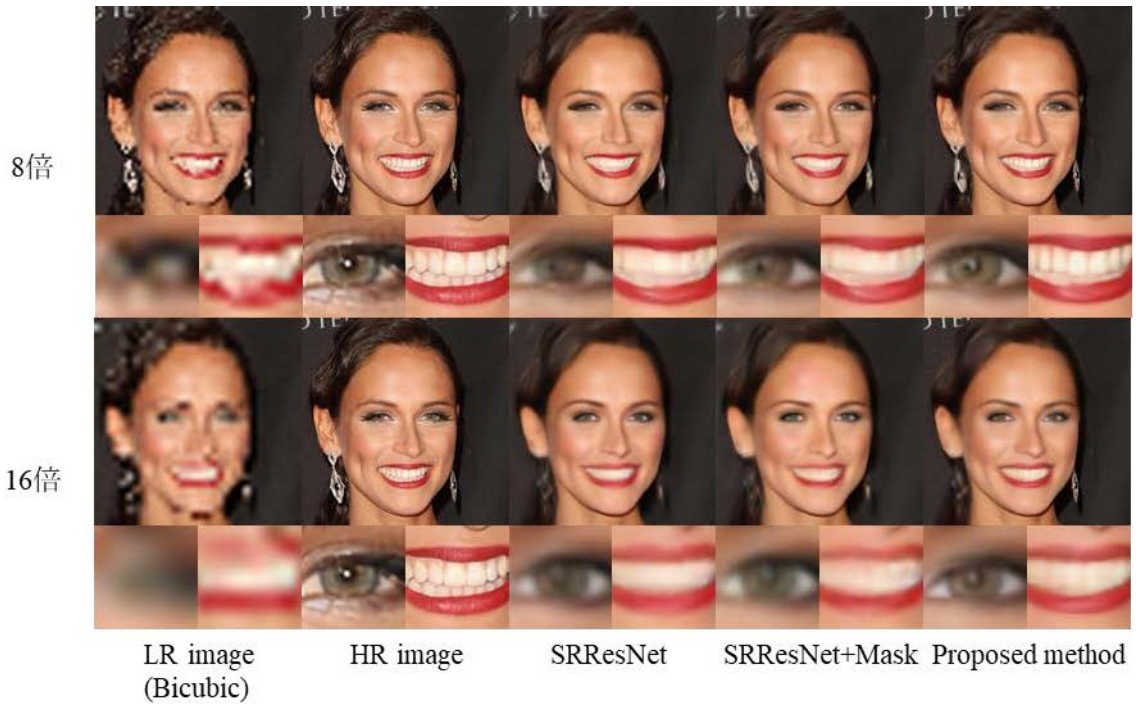


図 7 提案手法と従来手法による超解像画像の比較（上段：8 倍拡大画像，下段：16 倍拡大画像）

Fig. 7 Comparison of the super-resolution images with the proposed method and the conventional methods (upper: ×8, lower: ×16).

表 2 提案手法と従来手法による超解像画像の忠実度の比較（8 倍，16 倍拡大画像）

Table 2 Comparison of the fidelities of the super-resolution images with the proposed method and the conventional methods (×8, ×16).

Scale	Bicubic	SRResNet	SRResNet + Mask	**Proposed method**
×8	40.12±3.63 / 0.9988±0.0063	41.60±3.77 / 0.9991±0.0051	41.98±3.76 / 0.9990±0.0070	**42.42±3.58 / 0.9993±0.0039**
×16	36.45±3.38 / 0.9959±0.0173	38.77±3.41 / 0.9977±0.0108	38.69±3.37 / 0.9977±0.0106	**38.98±3.32 / 0.9978±0.0264**

PSNR: Ave.±Std.[dB] / SSIM: Ave.±Std.

顔画像の高解像度化において，4 倍だけでなく，8 倍，16 倍の超解像処理が望まれる状況もあると考える．そこで，**図 7** と**表 2** に，提案手法および SRResNet，SRResNet+Mask を 8 倍，16 倍の超解像処理に応用した結果を示す．ここでは，各ネットワークに顔画像（512×512 画素）と，1/8，1/16 のサイズにダウンサンプリングした低解像画像（64×64 画素，32×32 画素）の信号パターンの関係を学習させた．8 倍，16 倍ともに，提案手法は，SRResNet，SRResNet+Mask と比べ，パーツの境界をより鮮明に再構成することが確認できた．提案手法の平均 PSNR（8 倍：42.42 dB; 16 倍：38.98 dB）は，SRResNet（8 倍：41.60 dB，$p < 0.001$; 16 倍：38.77 dB，$p = 0.0035$），SRResNet+Mask（8 倍：41.98 dB，$p < 0.001$; 16 倍：38.69 dB，$p < 0.001$）より有意に高い結果が得られた．また，平均 SSIM においても，提案手法（8 倍：0.9993; 16 倍：0.9978）が，SRResNet（8 倍：0.9991，$p < 0.001$; 16 倍：0.9977，$p < 0.001$），SRResNet+Mask（8 倍：0.9990，$p < 0.001$; 16 倍：0.9977，$p < 0.001$）より有意に高かった．しかし，8 倍，16 倍の高解像度化において，提案手法を実用化するためには，さらなる改善が必要である．

今後の研究課題として，以下があげられる．

(1) 顔パーツのセグメンテーション法

本研究では，各顔パーツに手動でアノテーションが付与されたラベル画像を用いた．したがって，提案手法を顔画像に適用するためには，顔パーツのラベル画像が必要となる．我々の研究グループでは，顔パーツのセマンティックセグメンテーション法の開発にも取り組んでいる [25]．今後，提案手法とセマンティックセグメンテーション法を組み合わせることにより，ラベル画像を不要とするネットワークに改良する予定である．

(2) 主観的評価指標と一致する客観的評価指標の検討

本研究の観察者実験で，100 ペア中平均 92.33 ペアにおいて提案手法の超解像画像の知覚品質が高いと選択されたにもかかわらず，PSNR や SSIM に大きな差はなかった．したがって，PSNR や SSIM は主観的評価を反映できてい

ないと考える．主観的評価を反映した客観的評価指標を定義できれば，ネットワーク学習時の損失関数としても使用することが可能であり，より知覚品質の高い超解像画像が生成できると期待する．一方，本研究の観察者実験で採用した Two-alternative forced choice 法が高解像度化に対する主観的評価として，適切であったのかも検討する必要がある．そこで，これまで提案されている画質に関する客観的指標と主観的評価の関係を明らかにし，主観的評価をより反映する客観的指標の確立に取り組む．

(3) 高倍率の超解像処理への展開

図 7 と表 2 で示したように，8 倍，16 倍の高解像度化に応用するためには，提案手法をさらに改善する必要がある．そこで今後は，ネットワーク内部で，2 倍，4 倍，8 倍と段階的に高解像度化する手法 [26] を参考に，より高倍率の超解像処理に対応できるネットワークの改良に取り組む．

(4) 顔認証システムへの応用

近年，防犯/監視カメラ映像から，特定の人物を検索する顔認証システムが普及している [27]．しかし，防犯/監視カメラ映像の解像度が低く，十分な顔認証精度が得られないことがある．顔認証システムに提案手法を前処理として導入し，映像を高解像度化することにより，顔認証精度を向上できると期待する．今後，防犯/監視カメラ映像の顔認証システムにおける提案手法の有用性を検証する予定である．

6. まとめ

本研究では，各顔パーツを対象とした超解像 CNN モデルをそれぞれ用意したネットワークを構築し，各超解像 CNN モデルの出力を合成することにより，顔画像を高解像度化する手法を開発した．そして，提案手法と従来手法である SRResNet で生成した超解像画像に対し，主観的な知覚品質を評価するための観察者実験を行った．その結果，100 ペア中平均 92.33 ペアにおいて，提案手法の超解像画像が高画質と評価され，提案手法の有用性が示唆された．

参考文献

[1] Timofte, R., De Smet, V. and Gool, L.V.: A+: Adjusted Anchored Neighborhood Regression for Fast Super-Resolution, *Asian Conference on Computer Vision*, pp.111–126 (2014).

[2] Yang, J., Wright, J., Huang, T.S. and Ma, Y.: Image super-resolution via sparce representation, *IEEE Trans. Image processing*, pp.2861–2873 (2010).

[3] Dong, C., Loy, C.C., He, K. and Tang, X.: Image super-resolution using deep convolutional networks, *IEEE Trans. Pattern Analysis and Machine Intelligence*, Vol.38, No.2, pp.295–307 (2015).

[4] Dong, C., Loy, C.C. and Tang, X.: Accelerating the super-resolution convolutional neural network, *European Conference on Computer Vision*, pp.391–407 (2016).

[5] Kim, J., Lee, J.K. and Lee, K.M.: Accurate image super-resolution using very deep convolutional networks, *Proc. IEEE Conference on Computer Vision and Pattern Recognition*, pp.1646–1654 (2016).

[6] Ledig, C., Theis, L., Huszár, F., Caballero, F.J., Cunningham, A., Acosta, A. and Shi, W.: Photo-realistic single image super-resolution using a generative adversarial network, *Proc. IEEE Conference on Computer Vision and Pattern Recognition*, pp.4681–4690 (2017).

[7] Zhou, E., Fan, H., Cao, Z., Jian, Y. and Yin, Q.: Learning face hallucination in the Wild, *29th AAAI Conference on Artificial Intelligence* (2015).

[8] Chen, Y., Tai, Y., Liu, X., Shen, C. and Yang, J.: Fsrnet: End-to-end learning face super-resolution with facial priors, *Proc. IEEE Conference on Computer Vision and Pattern Recognition*, pp.2492–2501 (2018).

[9] Karalot, R., Li, T. and Porikli, F.: Component attention guided face super-resolution network, *Proc. IEEE/CVF Winter Conference on Applications of Computer Vision*, pp.370–380 (2020).

[10] Zhao, T. and Zhang, C.: Saan: Semantic attention adaptation network for face super-resolution, *Proc. IEEE International Conference on Multimedia and Expo*, pp.1–6 (2020).

[11] Ma, C., Jiang, Z., Rao, Y., Lu, J. and Zhou, J.: Deep face super-resolution with iterative collaboration between attentive recovery and landmark estimation, *Proc. IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp.5569–5578 (2020).

[12] Liu, Z., Luo, P., Wang, X. and Tang, X.: Deep learning face attributes in the wild, *Proc. IEEE International Conference on Computer Vision*, pp.3730–3738 (2015).

[13] He, K., Zhang, X., Ren, S. and Sun, J.: Deep residual learning for image recognition, *Proc. IEEE Conference on Computer Vision and Pattern Recognition*, pp.770–778 (2016).

[14] Loffe, S. and Szegedy, C.: Batch normalization: Accelerating deep network training by reducing internal covariate shift, *International Conference on Machine Learning*, pp.448–456 (2015).

[15] He, K., Zhang, X., Ren, S. and Sun, J.: Delving deep into rectifiers: Surpassing human-level performance on imagenet classification, *Proc. IEEE International Conference on Computer Vision*, pp.1026–1034 (2015).

[16] He, K., Zhang, X., Ren, S. and Sun, J.: Deep residual learning for image recognition, *Proc. IEEE Conference on Computer Vision and Pattern Recognition*, pp.770–778 (2016).

[17] He, K., Zhang, X., Ren, S. and Sun, J.: Identity mappings in deep residual networks, *European Conference on Computer Vision*, pp.630–645 (2016).

[18] Kim, J., Lee, J.K. and Lee, K.M.: Deeply-recursive convolutional network for image super-resolution, *Proc. IEEE Conference on Computer Vision and Pattern Recognition*, pp.1637–1645 (2016).

[19] Shi, W., Caballero, J., Huszár, F., Totz, J., Aitken, A.P., Bishop, R. and Wang, Z.: Real-time single image and video super-resolution using an efficient sub-pixel convolutional neural network, *Proc. IEEE Conference on Computer Vision and Pattern Recognition*, pp.1874–1883 (2016).

[20] Blau, Y. and Michaeli, T.: The perception-distortion tradeoff, *Proc. IEEE Conference on Computer Vision and Pattern Recognition*, pp.6228–6237 (2011).

[21] Kendall, M. and Gibbons, J.D.: *Rank correlation methods*, 5th edition, Oxford University Press (1990).

[22] Keys, R.: Cubic convolution interpolation for digital image processing, *IEEE Trans. Acoustics, Speech and Signal Processing*, pp.1153–1160 (1981).
[23] Johnson, J., Alahi, A. and Fei-Fei, L.: Perceptual losses for real-time style transfer and super-resolution, *European Conference on Computer Vision*, pp.694–711 (2016).
[24] Van der Maaten, L. and Hinton, G.: Visualizing data using t-SNE, *Journal of Machine Learning Research*, pp.2579–2605 (2008).
[25] 宮本 旭，檜作彰良，中山良平：Dice Loss と Multi Decoder Losses を用いた顔パーツのセマンティックセグメンテーション，第 19 回情報科学技術フォーラム（FIT2020）(2020).
[26] Zhang, Y., Wu, Y. and Chen, L.: MSFSR: A Multi-Stage Face Super-Resolution with Accurate Facial Representation via Enhanced Facial Boundaries, *Proc. IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops*, pp.2120–2129 (2020).
[27] 矢野光太郎，河合智明：監視カメラにおける映像認識の動向，日本画像学会誌，Vol.55, No.3, pp.341–347 (2016).

丸井 勇輝

2020 年立命館大学理工学部電子情報工学科卒業．同大学大学院理工学研究科電子システム専攻博士前期課程在籍中．

檜作 彰良

2014 年三重大学大学院工学研究科博士後期課程修了，博士（工学）．2014 年みずほ情報総研情報通信研究部入社．2018 年みずほ情報総研情報通信研究部退社．2018 年立命館大学理工学部電子情報工学科助教，現在に至る．電子情報通信学会，電気学会，医学物理学会，日本医用画像工学会会員．

中山 良平 （正会員）

1999 年宮崎大学工学部情報工学科卒業．2001 年同大学大学院工学研究科修士課程修了．2005 年三重大学大学院医学系研究科博士課程修了．同年三重大学医学部附属病院助教，2015 年立命館大学理工学部准教授，2020 年同教授．この間，2008 年シカゴ大学放射線科 visiting assistant professor．主に医用画像を対象とした画像認識，機械学習に関する研究に従事．博士（工学），博士（医学）．IEEE，SPIE，電子情報通信学会，日本医用画像工学会，日本画像情報学会，日本放射線技術学会各会員．