Vol.2019-CVIM-217 No.29
2019/5/30

直接的知覚損失関数と間接的知覚損失関数を用いた画像超解像の検証

吉田 智樹[1,a)]　秋田 和俊[1,b)]　Muhammad Haris[1,c)]　Greg Shakhnarovich[2]　浮田 宗伯[1,d)]

概要：超解像（SR）は，低解像度画像（LR 画像）を高解像度画像（HR 画像）に変換する技術であり、近年は深層学習によってその精度を向上させている．従来の教師あり学習に基づく超解像手法と同様に，超解像モデルはその出力 SR 画像が可能な限りピクセル単位で HR 画像と等しくなるように学習されることが多く， SR 画像と HR 画像との復元誤差は，これら 2 画像間の平均二乗誤差（MSE）によって定量化され，損失関数に利用されている．しかし，近年の研究で人が見て綺麗と感じる（つまり知覚的精度が良い）SR 画像の生成には超解像モデルに損失関数に MSE のみを用いるだけでは不十分であることが明らかにされた．そこで，知覚的精度が良い SR 画像を生成するために，様々なコンピュータビジョンの問題において広く使用されている敵対的損失関数（GAN 損失と呼ぶことにする）のような数種類の損失関数の利用が提案され，それらの有用性が実証されている．これらの損失関数は知覚的精度に良い影響を与えることができる一方で，どの画質評価値が向上したのかなどが完全には明らかにされていない．そこで本研究ではこのような間接的に知覚的精度に影響を与えるのではなく，知覚的精度を直接的に改善する代替損失関数として直接的知覚損失関数を提案する．本研究では 2 つの直接的知覚損失関数である PCA 損失と NIQE 損失を提案し，具体的に効力が明らかになっていない先述の損失関数とともに直接的知覚損失関数の効力を検証した．この検証により，PCA 損失は MSE 損失のみで学習した結果と比較し，復元誤差を維持しながら知覚的精度を改善するのに有用であることが分かった．

キーワード：超解像，損失関数，主成分分析（PCA）

1. はじめに

画像超解像技術とは Irani ら [1] に始まる，低解像画像を高解像画像に変換・補間する技術である．この超解像技術は近年，機械学習・深層学習を応用しており，それらの進化に伴い精度を高めている．特に Krizhevsky ら [2] が発表した畳み込みニューラルネットワーク（CNN）の応用により飛躍的に精度を上げ，CNN を利用した Chao ら [3] の super-resolution convolutional neural networks（SRCNN）は，ピーク信号対雑音比（PSNR）や平均二乗誤差（MSE）といった一般的な画像評価指標で CNN 以前の超解像モデルよりも良い成果を上げた．この CNN を用いた超解像モデルは，その後も主に PSNR や MSE などの復元誤差に関する誤差指標において良好な結果を更新し続けていた．

このような超解像モデルの学習において，復元誤差評価値が良い SR 画像を生成するモデルは，人が綺麗と感じる（知覚的に良い）SR 画像を生成できると思われていたため，多くの超解像ネットワークの学習で用いられる損失関数には主に MSE が用いられていた．しかし，様々な超解像ネットワークや手法が提案される中で，多くの研究者により復元誤差のみを損失関数にして生成された SR 画像は，知覚的精度がそれほど高くないという事が指摘された．後に復元誤差と知覚的精度の関係は，Blau ら [9], [10] によって数値的にもトレードオフの関係にあることが示された．

そこで知覚的精度を向上させるために，Johonson ら [19] は Simonyan ら [7] の VGG を用いて，特徴量空間における生成画像の特徴量と HR 画像の特徴量の距離を測る perceptual 損失（本研究では VGG 損失と呼ぶことにする）を提案し，Gatys ら [15], [16] は VGG によって得られる特徴量のグラム行列の MSE を測る style 損失を提案した．また，Ledig ら [6] は Goodfellow ら [5] の敵対的生成ネットワーク（GAN）を用いた generative adversarial network

[1] 豊田工業大学
Toyota Technological Institute
[2] 豊田工業大学シカゴ校
Toyota Technological Institute at Chicago
[a)] sd19455@toyota-ti.ac.jp
[b)] sd19401@toyota-ti.ac.jp
[c)] mharis@toyota-ti.ac.jp
[d)] ukita@toyota-ti.ac.jp

図 1 HR 画像，LR 画像（バイキュービック補間），復元誤差 MSE 損失のみで生成された SR 画像 SR1 と知覚評価を考慮した損失関数を用いて生成された SR 画像 SR2 の比較．SR1 の輪郭は不鮮明であるが，SR2 の人物の輪郭は先鋭的ではっきりしている．

for image super-resolution（SRGAN）を提案し，そこで用いられる損失関数（本研究では GAN 損失と呼ぶ）で知覚的に良い画像を生成しようとした．これらの損失関数の有用性は様々な超解像ネットワークによって示されており，例えば style 損失は Sajjiadi ら [4] の EnhanceNet で，VGG 損失と GAN 損失は SRGAN で示されている．

ここで，我々は VGG 損失や style 損失，GAN 損失といった損失関数は，結果として知覚的精度を高めているに過ぎず，知覚的精度に関する値を直接最適化しているわけではないことに注目した．このことから，本研究では上記の 3 つのような損失関数を間接的知覚損失関数と呼ぶことにする．そこで，本研究では人の知覚的画質評価を直接損失関数に取り入れた直接的知覚損失関数を提案し，知覚的精度に関する値を直接最適化することを狙う．つまり，人の感覚を機械学習に盛り込み，自動的に知覚的に良い SR 画像を生成することを目指す．本研究では，損失関数に実装する知覚的画質評価として，競技会 [10] における知覚的評価指標 Pi（Blau ら [10] は Perceptual index と表現）を参考にした．そして，Pi を構成する NIQE [8] および Ma [13] の一部である主成分分析（PCA）を損失関数として実装し，それらの有用性を検証する実験を行った．

2. 関連研究

本節では，検証実験で用いる間接的知覚損失関数と直接的損失関数として実装する画像評価指標について述べる．

2.1 間接的知覚損失関数

- VGG 損失

本研究で VGG 損失と呼ぶ損失関数は Johnson ら [19] によって提案された損失関数 perceptual 損失である．VGG 損失提案当時から Dosovitskiy ら [20] などによって，画像分類などを目的に学習された CNN に画像を入力したときに抽出される特徴量を用いることが，高画質な SR 画像生成に有用であるということが指摘されていた．この知見から Johnson らは，Simonyan ら [7] が提案した VGG16（ImageNet データセット [14] で学習済み）を使用して得られる特徴量を用いた損失関数を提案した．具体的には，SR 画像と HR 画像をそれぞれ VGG16 に入力したときに得られる特徴量行列の距離の二乗を，特徴量行列の行数と列数，チャンネル数の三つの積で割って求めている．これにより，特徴量空間に写像された SR 画像の特徴量と HR 画像の特徴量の差を最小化して画像に内在する特徴量の距離を小さくし，知覚的精度のよい SR 画像を生成しようとした．本研究では，後に Gatys ら [15] によって有用とされた VGG19 を用いる．VGG19 は Johnson ら [19] が使っていた VGG16 よりも畳み込み層が多く，その分高レベルな特徴量を抽出できるためである．用いる特徴量は Haris ら [11] に倣い，VGG19 の 9 層目と 18 層目で得られる特徴量に対して上記の演算を行い，9 層目と 18 層目の演算結果を足し合わせる．中間層と出力層に近い層から特徴量を抽出することで，高レベルかつ比較的局所的な特徴量と広域的な特徴量の両方を得られる．よって，VGG 損失 L_{VGG} を次のようにした．

$$L_{VGG} = \sum_{k=9,18} \frac{1}{CHW} \sum_{i,j} (f_{SR_k}(i,j) - f_{HR_k}(i,j))^2) \quad (1)$$

$f_{SR_k}(i,j)$ は SR 画像を VGG19 に入力し，VGG の k 層目の出力である特徴量行列の i 行 j 列目の値，$f_{HR_k}(i,j)$ は HR 画像を VGG19 に入力し，VGG の k 層目の出力である特徴量行列の i 行 j 列目の値である．C，H，W はそれぞれチャネル数，画像の高さと幅である．

- style 損失

次に，style 損失について述べる．style 損失は元々，Gatys ら [15], [16] が画風変換に用いた損失関数として提案したものであるが，Sajidi ら [4] は超解像モデル EnhanceNet の学習にも利用し成果を上げることができたことで，超解像モデルでも有用であることが分かった．Sajidi らの style 損失は VGG 損失と同様に，まず SR 画像と HR 画像をそれぞれ学習済みの VGG ネットワークに入力し，任意の畳み込み層から特徴量行列を得る．得られた特徴量行列からグラム行列を導出し，グラム行列の距離の二乗を算出する．グラム行列は特徴量行列と特徴量行列の共役転置行列をかけることで得られる．この処理を複数の畳み込み層の特徴量行列に対して行うことで得られる値を足し合わせた値を style 損失としている．Gatys によれば，グラム行列から画像内の定常的な特徴量を得られるとしているため，Sajidi らは超解像に用いて HR 画像内の描写を SR 画像にも生成できるとしている．本研究では VGG 損失と同様に，Haris ら [11] に従い，SR 画像を VGG19 に入力したときの 9 層目と 18 層目から得られる特徴量のグラム行列と，同様にして HR 画像から得られるグラム行列の MSE を取り，それらの和を取ることで style 損失を計算する．よって style 損失 L_{style} は次のようにした．

$$L_{style} = \sum_{k=9,18} \frac{1}{CHW} \sum_{i,j} (\mathbf{g_{SR_k}}(i,j) - \mathbf{g_{HR_k}}(i,j))^2 \quad (2)$$

$\mathbf{g_{SR_k}}(i,j)$ は SR 画像を VGG19 に入力し，VGG の k 層目の出力である特徴量行列のグラム行列の i 行 j 列目の値，$\mathbf{g_{HR_k}}(i,j)$ は HR 画像を VGG19 に入力し，VGG の k 層目の出力である特徴量行列のグラム行列の i 行 j 列目の値である．C，H，W はそれぞれチャネル数，画像の高さと幅である．

- GAN 損失

generative adversarial nets（GAN）とは，Goodfellow ら [5] によって提案され，現在画像生成分野で広く使われている深層学習モデルである．GAN では画像生成器のほかに画像識別器を用い，その識別器の判断を用いることでより目的に適した画像を生成することを目的としている．この識別器の識別結果を生成器の損失関数に加えて学習させることで，識別器でも区別できないほどの画像を生成できる生成器を作成していく．この GAN の目的関数は Goodfellow ら [5] によって次のように定義されている．

$$\min_G \max_D V(D,G) = E_{\mathbf{x}} \sim p_{data}(\mathbf{x})[\log D(\mathbf{x})] + E_{\mathbf{z}\sim p_z(\mathbf{z})}[\log(1 - D(G(\mathbf{z})))] \quad (3)$$

$\mathbf{x}$ は識別器への入力，$D(\mathbf{x})$ は識別器の出力（確率値）であり，$\mathbf{z}$ は生成器への入力で，$G(\mathbf{z})$ は生成器の出力画像を表す．本研究で GAN 損失 L_{GAN} と呼ぶ損失関数は，式（3）の第 2 項に相当する関数としており，上記の識別器に SR 画像 $\mathbf{I}$ を入力したときの出力と，「教師データである」というラベル（$t = 1$）を引数として次式のようにした．

$$L_{GAN} = -t \log D(\mathbf{I}) = -\log D(\mathbf{I}) \quad (4)$$

本研究では Haris ら [11] に倣い，SRGAN[6] と同様の構造を持つ識別器を用いる．

2.2 知覚的画質評価

本研究で参考にした人の知覚的画質評価は，Blau ら [9], [10] の Pi（Blau ら [10] は Perceptual index と表現）であり，二つの評価指標の Ma と NIQE の評価値によって構成されている．本節ではこれら Ma と NIQE ついて紹介する．

画像の画質評価法は大きく，全参照モデルと非参照モデルの 2 つに分類できる．この内，全参照モデルの多くは人間が評価した画像データベースの存在が前提となっている．全参照モデルは，その画像データベースから事前に「人間らしい評価方法」を学習し，そのモデルを使って画質を評価するものである．今回参考にした Blau ら [10] の Pi に用いられる Ma らの評価指標 [13] は，この事前学習による全参照評価モデルである．

Ma らの研究では，まず 9 種の超解像手法で生成された膨大な SR 画像を 50 人の被験者に評価させた．その結果から，人間の主観評価を再現するために局所周波数特徴量，大域周波数特徴量，空間的特徴量の 3 つが重要と解析し，それぞれの特徴量を離散コサイン変換・ウェーブレット分解・主成分分析（PCA）によって得ることを提案した．そして，それらの特徴量をランダムフォレストに入力し，出力値を被験者実験で得られた評価値となるようにランダムフォレストを学習させ，評価モデルを生成した．よって Ma では画像から先述の 3 つの特徴量を抽出し，それらをランダムフォレストに入力し，得られる出力値の重み付き線形和を評価値としている．

一方の NIQE は，Ma と違い事前学習を行わない非参照モデル評価基準である．NIQE では，人間による画像評価値を使わない代わりに，予め人が自然と感じる画像を 125 枚集め，その画像群（画像コーパス）から統計的規則性に従って得られる特徴量のみを使う．この特徴量を Mittal ら [8] は natural scene static 特徴量（NSS 特徴量）と呼んでいる．NIQE の評価方法は，評価対象画像から得られる NSS 特徴量を多変数ガウスモデルに当てはめる．この分布と画像コーパスから得られる NSS 特徴量を多変数ガウスモデルに当てはめたときの分布の距離を評価値としており，この距離が小さいほど対象画像はより自然で，知覚的に良いとしている．[8] らの実験では，NIQE による評価は，参照型評価モデル以上の評価精度で，かつ，事前学習を要する非参照モデルと同等もしくは同等以上の評価精度を示すことができている．

Ma の評価法は，ランダムフォレストへの入力前の値や行列の導出は PCA などの広く知られた演算により行われており，NIQE は人の知覚的画質評価により近い評価精度を誇り，かつ，データベースの学習を必要としない統計的規則から得られる値のみで評価を行う．以上の観点から，機械学習の損失関数に取り入れやすいと考え，本研究ではこれら 2 つ指標を損失関数とした直接的知覚損失関数を実装した．

3. 提案手法

3.1 直接的知覚損失関数の実装

本研究では，人の知覚的画質評価を損失関数に実装した直接的知覚損失関数を提案し，これにより知覚的精度に関する値を直接最適化することを狙う．この直接的知覚損失関数を用いることで，人の感覚を機械学習に反映させ，自動的に知覚的精度の良い SR 画像を生成することを目指す．

本研究では，2 章で挙げた Ma と NIQE を元に，2 つの直接的知覚損失関数を実装した．

Ma の評価に関しては，3 つの特徴量抽出処理の内 PCA のみを損失関数として実装した．以下では，これを PCA 損失と呼ぶことにする．これは，PCA の処理に要する時間が学習中で用いる損失関数として適した短さであり，かつ，Ma の評価値への寄与率が高いためである．Ma らは画像に PCA を行うことで画像内の空間的不連続性が取得され，これが知覚的精度にとって重要であるとしている．

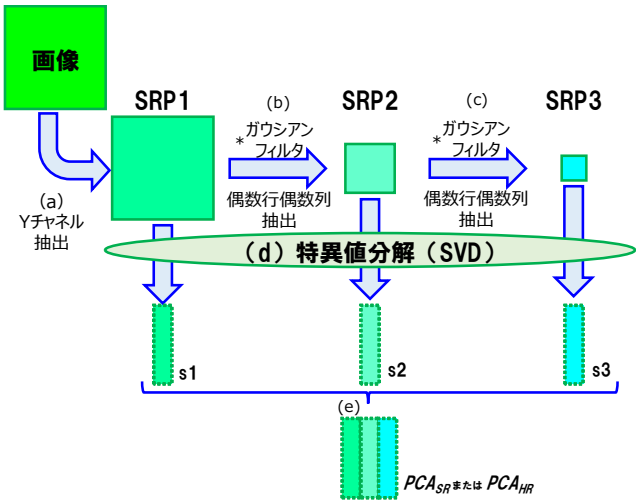
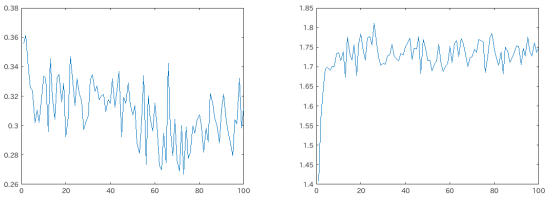


図 2 PCA_{SR} または PCA_{HR} を得る概略図．本図中で用いるガウシアンフィルタは全て 3×3 である．(a) で輝度を抽出して SRP1 を得る．(b)，(c) により SRP2，SRP3 を得る．(d) で SRP1，2，3 の特異値行列の対角成分 s1，s2，s3 を得る．(e) でこれらを列方向に結合し，PCA_{SR} または PCA_{HR} を得る．



(a) 各エポックの PCA_{SR} と PCA_{HR} の MSE　　(b) 各エポックの Ma の PCA に基づく評価

図 3 PCA 損失を用いた学習で (a) は縦軸を PCA_{SR} と PCA_{HR} の MSE とした図，(b) は縦軸を Ma の評価内の PCA に基づく評価値とした図（いずれも横軸はエポック数であり，100 エポックの学習を行った）．おおよそ PCA_{SR} と PCA_{HR} の MSE が小さいと Ma の PCA に基づく評価値は大きいという相関が見られる．

次に，本研究で実装した Ma の PCA 損失の計算について述べる．PCA 損失の計算法は図 2 で示す．

まず，図 2 中（a）に示したように，生成された SR 画像の輝度である Y チャネル成分 SRP1 を得る．次に図 2 中（b）のように Y チャネル成分 SRP1 にガウシアンフィルタ（サイズは 3×3）を掛け，偶数行偶数列のピクセルのみ抽出して SRP2 を用意する．同様に図 2 中（c）のように SRP2 にガウシアンフィルタを掛け，偶数行偶数列ピクセルを抽出した SRP3 を用意する．これらの SRP1，SRP2，SRP3 それぞれに図 2 中（d）で特異値分解（SVD）を行い，それぞれの特異値行列の対角成分（特異値ベクトル）s1，s2，s3 を得る．

以上のようにして得られた特異値ベクトルを列方向に結合させ，図 2 中（e）で 25 行 3 列の特徴量行列を得る（PCA_{SR}）．この処理を HR 画像に対しても行い，PCA_{HR} を得る．これら PCA_{SR} と PCA_{HR} の MSE を取り PCA 損失とした．

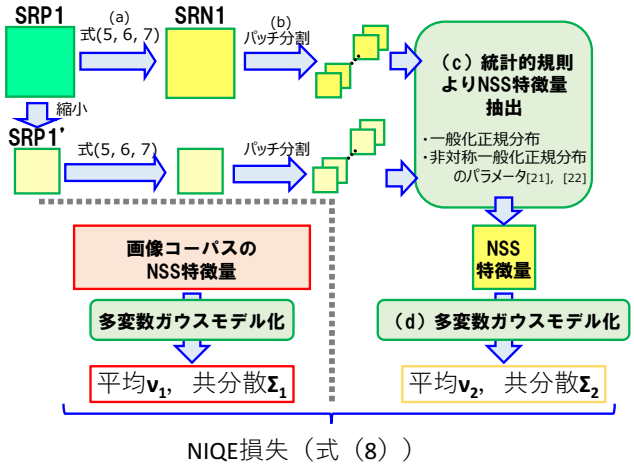


図 4 NIQE 損失の計算処理の概略図．SRP1 は図 2 の SRP1 と同じである．(a) で輝度に基づき正規化を行い，(b) でパッチ分割を行う．(c) で統計的規則に基づき SRN1 の NSS 特徴量行列を得る．SRP1 をバイキュービック補間により辺々 2 分の 1 に縮小した SRP1 ' に対しても同様の計算により SRP1 ' の NSS 特徴量を得る．2 つの NSS 特徴量行列を列方向に結合させた後，(d) で多変数ガウスモデルに当てはめ，そのガウスモデルの平均行列と共分散行列を得る．自然画像コーパスから得られる平均行列と共分散行列は [8] が公開しているものを用い，これらを式 8 に当てはめて NIQE 損失を得る．

上記 Ma の PCA 処理では，SR 画像の Y チャネル成分 SRP1 のみではなく，SRP2 や SRP3 からも特異値ベクトルを得ている．これは，生成された SR 画像の大きさによらず，超解像によって補間される画像描写を評価できるようにするためであり，データ拡張の意味合いも含まれている．

また本研究の PCA 損失の実装では，Ma ら [13] が評価のために用いたランダムフォレストは実装していない．これは，PCA 損失を用いた学習で図 3 のように PCA_{SR} と PCA_{HR} の MSE の値と Ma で評価した際の PCA に基づく評価（特異値行列をランダムフォレストに入力した時の出力値）を各エポックで確認してみたところ，PCA_{SR} と PCA_{HR} の MSE が小さいと，Ma の評価内の PCA による評価値は大きくなったためである．つまり，ランダムフォレストの入力となる PCA_{SR} と PCA_{HR} の差分は，知覚評価値との相関が高いため，そのまま評価値として利用可能であると考えられる．そこで損失関数としての計算時間と計算コストを可能な限り抑えるため，PCA 損失の定義を PCA_{SR} と PCA_{HR} の平均二乗誤差とした．

NIQE の評価に関しては，本研究では [21] を参考にしつつ評価に要する処理で得られる画質評価値を NIQE 損失とした．図 4 には，NIQE 損失の処理の流れの概略図を示した．

NIQE は画像の輝度を正規化（標準化）した行列を元に評価処理を行うため，NIQE 損失でもまずは，PCA 損失のように SRP1 を用意する．次に図 4 中（a）の処理で，式（5）のように SRP1 にサイズが 7×7 のガウシアンフィルタを掛けて行列 μ を得る．次に，式（6）のように SRP1 と μ の画素ごとの差の二乗にサイズ 7×7 のガウシアンフィルタを掛け，それらの平方根をとることで行列 σ を得る．これら μ，σ をそれぞれ SR 画像の平均と標準偏差として用い，次式のようにして標準化された SRN1 の各画素値を求める．

$$\mu(i,j) = \sum_{k=-3}^{3}\sum_{l=-3}^{3} w_{k,l} SRP1(i+k, j+l) \tag{5}$$

$$\sigma(i,j) = \sqrt{\sum_{k=-3}^{3}\sum_{l=-3}^{3} w_{k,l}[SRP1(i+k, j+l) - \mu(i,j)]^2} \tag{6}$$

$$SRN1(i,j) = \frac{SRP1(i,j) - \mu(i,j)}{\sigma(i,j) + 1} \tag{7}$$

次に図 4 中（b）のように SRN1 を 96×96 のミニパッチに分割し，各々のパッチについて図 4 中（c）で以下の 2 つの分布を用いて，NSS 特徴量を取得する．1 つ目の分布は一般化正規分布であり，2 つ目の分布は非対称一般化正規分布である．2 つの分布は共にミニパッチから求める．1 つ目の一般化正規分布については，ミニパッチ内の各画素から平均を 0 とした分布を求める．また，一般化正規分布のパラメータと非対称一般化正規分布のパラメータは Mittal らと同様に [18] が提案した手法によって得る．ここで得られるパラメータが NSS 特徴量であり，この特徴量抽出をミニパッチ毎に行う．NSS 特徴量は SRP1 をバイキュービック補間により辺々を半分に縮小させた SRP1 ’ からも得る．SRP1 と SRP1’ から得られた全ての NSS 特徴量を列方向に結合させベクトル化した後，多変数ガウスモデルに当てはめ（図 4 中（d））, 多変数ガウスモデルの平均行列 ν_2 と共分散行列 $\mathbf{\Sigma}_2$ を得る．画像コーパスについても事前に同様の処理をし、それらから得られる 2 つの行列である平均行列 ν_1 と共分散行列 $\mathbf{\Sigma}_1$ を予め用意しておく．この平均行列 ν_1 と共分散行列 $\mathbf{\Sigma}_1$ は [8] によって公開されているため本研究ではそれらを用いた．最後に，得られた 2 つの平均行列と 2 つの共分散行列について次の式（8）のように距離を取り，それを NIQE 損失 L_{NIQE} とした．

$$L_{NIQE} = \sqrt{(\nu_1 - \nu_2)^T \left(\frac{\Sigma_1 + \Sigma_2}{2}\right)^{-1} (\nu_1 - \nu_2)} \tag{8}$$

ただし，ν_1 と Σ_1 は画像コーパスから得られる平均行列と共分散行列，ν_2 と Σ_2 は SR 画像から得られる平均行列と共分散行列である．

NIQE では画像コーパスから多変数ガウスモデルを得るときに，画像をパッチ分割したものの内，高周波成分を多く含むパッチを選択し，それらの NSS 特徴量を用いて作成していた．この選択は Mittal らも参考にした Hassen ら [12] が述べた，人間が画像のシャープな領域（高周波成分を含む領域）から画質評価をする傾向があるという知見に基づいて行っている．このため NSS 特徴量の導出はパッチ毎に行う．

また，NIQE 損失の導出ではパッチの選択は行っていない．評価対象画像である SR 画像に対してもパッチの選択を行ってしまうと，画像の歪みを表す高周波領域の減少を正確に測れなくなるためである．

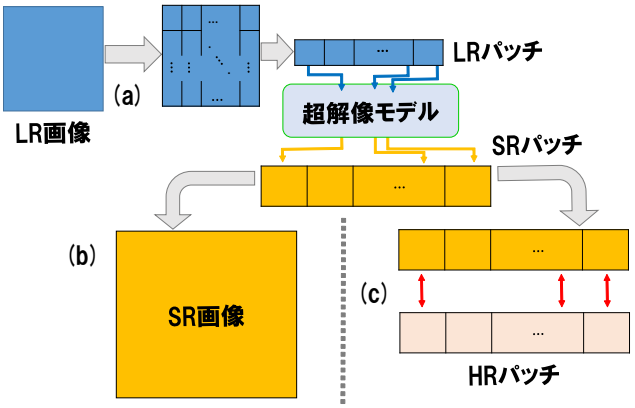


図 5 本研究における直接的知覚損失関数と間接的知覚損失関数，MSE 損失の計算法の概略図．（a）で LR 画像を超解像モデルの入力に適した LR パッチに分割して，超解像モデルに入力する．得られた SR パッチに対し，（b）で元の画像のように結合し直してから直接的知覚損失関数を計算し，間接的知覚損失関数と MSE 損失は（c）のようにパッチ状態のまま各々で計算を行う．

3.2 直接的知覚損失関数の計算法

Ma や NIQE は，画像全体を入力して 1 つの評価値を返すため，直接的知覚損失関数も同様にして，画像の広範囲に損失関数を適応することで本来の知覚的画質評価を学習に反映できるようになると考える．しかし，深層学習を用いた超解像モデルは，メモリの消費量を考慮して，画像の一部の小領域のみを用いて学習することがある．本研究で用いる DBPN も画像の一部の小領域を用いて学習を行う．この両者を整合させるために図 5 のような学習を行うようにした．

はじめに，図中（a）のように LR 画像を超解像ネットワークに適したサイズに分割して，LR パッチを得る．次に各々の LR パッチを超解像モデルに入力し，SR パッチを得る．直接的知覚損失関数の計算は図中（b）のように SR パッチを画像の形に結合した後に計算を行う．これにより，画質評価を再現できるようにする．

直接的知覚損失関数以外の計算は SR 画像と HR 画像間の画素毎に対する処理と同等であるため，図中（c）のようにパッチ状態のまま計算を行う．パッチごとに計算した後は，得られた値の平均を取るなどして，代表値を損失関数

につき 1 つ得る．最後に各損失関数から得た値の重み付き線形和を学習で用いる損失関数とし，学習を行う．この損失関数は，その他損失関数と区別するため，G 損失関数と呼ぶ．

4. 検証実験

4.1 実験条件

本研究では，Haris ら [11] の deep back-projection networks（DBPN）の内 dense-DBPN（D-DBPN）を超解像モデルとして用いる．この D-DBPN は競技会 [10] で高い知覚的精度をもつ SR 画像を生成できることが示されている．また，D-DBPN の出した結果は，本研究で用いる間接的知覚損失関数（VGG 損失，style 損失，GAN 損失）を用いていた．本研究では，LR 画像を辺々 4 倍にした SR 画像を生成することを D-DBPN に学習させた．

用いる損失関数は，直接的知覚損失関数として上記で提案した PCA 損失と NIQE 損失を，間接的知覚損失関数として，2 章で挙げた VGG 損失，style 損失，GAN 損失を用いる．また，本実験では知覚的精度の高い SR 画像を生成するための検証として，G 損失を MSE 損失のみの場合と MSE 損失と 1 つの知覚損失関数の重み付き線形和にした場合の超解像ネットワークの学習を行い，生成される SR 画像の定量的評価と定性的評価を行った．これは Blau ら [9], [10] が述べていたように，知覚的精度と復元誤差はトレードオフの関係であるため知覚的損失関数のみを G 損失とすると HR 画像と全く異なる SR 画像が生成される可能性があるためである．以上のようにして G 損失を構成し，学習を行うことで上記の知覚損失関数の効力を調査する．

まず，比較用に G 損失を MSE 損失のみにして D-DBPN を学習させた．この学習エポック数は 500 であり，これは 500 エポックで十分に G 損失が収束していたためである．次に，G 損失を MSE 損失と 5 つの知覚損失関数の内 1 つを用いて構成し学習させる．この 5 パターンの学習は MSE 損失のみで学習済みのモデルからの再学習である．

LR 画像のサイズは，Haris ら [11] に倣い 96 ピクセル ×96 ピクセルとし，LR パッチのサイズは 24 ピクセル ×24 ピクセルとした．学習率は 0.0001，バッチサイズを 2，最大学習エポック数を 500，最適化手法を Adam（α=0.9, β=0.99）とし，学習時のデータセットには [17] らの DIV2K の学習データセットを用いて行った．データ拡張は左右・上下反転や回転をランダムで行っている．また，学習モデルの評価のために用いたデータセットは競技会 [10] がテストデータとして公開したデータ（画像 100 枚）を用いた．

4.2 MSE 損失と知覚損失関数の重み探索法

D-DBPN において,G 損失計算時の MSE 損失と知覚損失関数の重みの比率選択は，次のようにした．まず，10 エポックの学習の中で各損失関数の最大値と最小値を求める．次に，1 エポック目の損失関数の値と 10 エポックの学習における最小の損失関数の値の差を最大値で割って正規化した減少度を得る．この減少度を MSE 損失と知覚損失関数でそれぞれ求め，減少度の和を求める．この減少度の和が最大のものを重みの組み合わせとした．

このように各損失関数の値を詳しく追うことで，どの重みによる学習が安定するのか，またどの組み合わせが復元誤差を向上させる MSE 損失に対して効果的に知覚損失関数の影響を与えられるのかを把握するためである．また，学習エポック数を 10 にしたのは，重みの比率が学習に適した場合，10 エポック程度の学習で損失関数の値は安定して下がることが分かったためである．

この重み比率の探索では，MSE 損失の重み係数を固定し，MSE 損失に対する間接的知覚損失関数の重みの比率は競技会 [10] で使われた D-DBPN の結果を参考にした．

表 1 MSE 損失に知覚損失関数を 1 つ加えて学習し，生成された SR 画像 100 枚を知覚的評価指標 Pi，復元誤差評価 RMSE, PSNR，SSIM で評価した平均値．Pi と RMSE は値が小さいほど精度が良く，PSNR と SSIM は値が大きいほど精度が良いとされる．赤字の MSE 損失に PCA 損失を加えた学習結果は青字の MSE 損失のみの学習結果と比べ, 復元誤差は保ちつつ，知覚的精度が大きく向上していることが分かる．

G 損失	Pi	RMSE	SSIM	PSNR
MSE(比較)	5.15	11.2	0.771	28.3
MSE+VGG	5.27	11.6	0.751	27.8
MSE+GAN	5.20	11.3	0.765	28.2
MSE+style	**5.02**	**11.3**	**0.768**	**28.2**
MSE+NIQE	**4.96**	**11.3**	**0.762**	**28.1**
MSE+PCA	**3.82**	**12.0**	**0.752**	**27.6**

4.3 実験結果

以下に，検証実験を行った結果を示す．学習したモデルから生成された SR 画像の定量的評価について，競技会 [10] の Pi により知覚的精度を評価し，ピーク信号対雑音比（PSNR），SSIM，二乗平均平方根誤差（RMSE）で復元誤差を評価する．Pi は値が小さいほど知覚的精度が良いという評価となり，表 1 のようになった．表 1 の青字は MSE 損失のみで学習したモデルによる SR 画像の評価である．

現段階で MSE 損失に VGG 損失あるいは GAN 損失を加えた学習結果は，MSE 損失のみで学習した結果と同様になり，効力を把握することはできなかった．この原因は，重み付き線形和によって G 損失を算出するときの VGG 損失の重み係数や GAN 損失の重み係数が不適切であったことや，あるいは計算方法が不適切であったことなどが考えられる．特に GAN 損失の計算方法については，識別器はミニパッチに分割した画像ではなく画像全体，あるいは本研究でのミニパッチよりも大きいサイズのパッチで識別す

図 6 MSE 損失，MSE 損失に style 損失，NIQE 損失，PCA 損失のいずれか 1 つを加えて学習して生成した 4 つの SR 画像と HR 画像，LR 画像（バイキュービック補間により生成）と各 SR 画像の Pi，PSNR，SSIM の値である．MSE 損失のみの学習と比較し，MSE 損失に PCA 損失を加えたものは画層中央下部の文字プレート内の文字の輪郭が先鋭化されている．

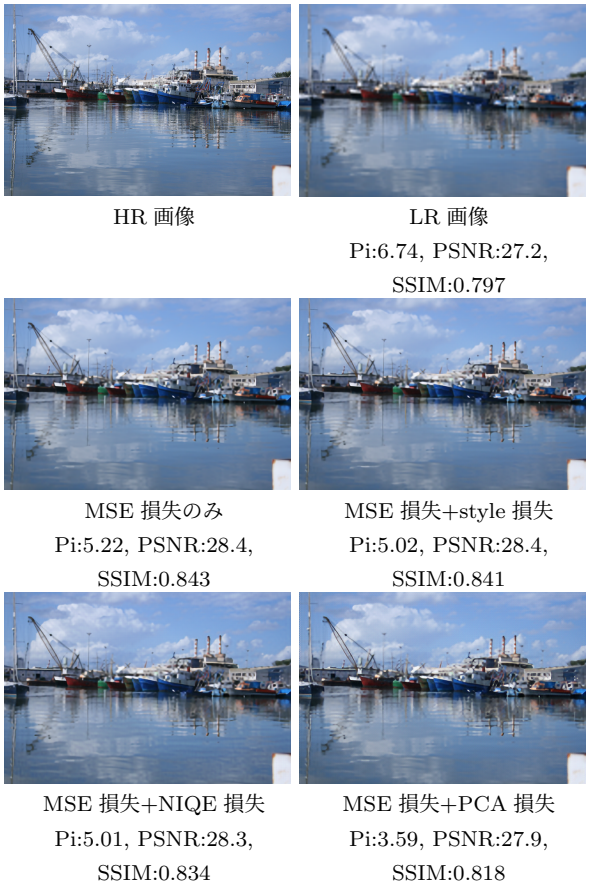


図 7 図 6 と同様にして生成された SR 画像と HR 画像，LR 画像．MSE 損失に NIQE 損失を加えて学習し，生成された SR 画像と各 SR 画像の Pi，PSNR，SSIM の値である．MSE 損失に NIQE 損失を加えた SR 画像は，画像右下の海の波紋が強調されるような画像となっている．

るため，直接的知覚損失関数と同様にして計算したほうが良いと考えられる．この場合は，計算に用いる使用メモリを節約するために，更なる工夫が必要である．

一方で，太字の style 損失，NIQE 損失，PCA 損失は MSE 損失のみで学習した結果と比べ，復元誤差をほぼ保ちながら知覚的精度を向上させることができるということが分かった．特に赤字の PCA 損失を加えた結果は，RMSE が 0.8 悪くなった代わりに Pi を 1.33 向上させることができている．競技会 [10] において損失関数に MSE 損失，VGG 損失，style 損失，GAN 損失を用いた D-DBPN は，RMSE が 11.46 で Pi が 2.938，RMSE が 12.40 で Pi が 2.199 であり RMSE が 0.94 悪くなる代わりに Pi が 0.739 向上したことと，Blau ら [9], [10] の SR 画像の知覚的精度と復元誤差がトレードオフの関係にあるという指摘を踏まえると，復元誤差をほぼ保ちつつ知覚的精度を大きく向上させた PCA 損失は有用であると考えられる．

次に，MSE 損失のみの学習結果よりも知覚的精度 Pi が良かった MSE 損失と style 損失，MSE 損失と NIQE 損失，MSE 損失と PCA 損失で学習した D-DBPN が生成した SR 画像と，MSE 損失のみで学習した D-DBPN が生成した SR 画像を見比べて定性的な評価を行う．図 6，7 には上記の 5 種の SR 画像から切り抜いた画像の一部と対応する HR 画像，LR 画像（バイキュービック補間により拡大）を示す．各損失関数の下の 4 つの数値はそれぞれ左から Pi，RMSE，PSNR，SSIM の値である．実際に画像を見て比較すると，MSE 損失に style 損失を加えた学習結果から得られる SR 画像と MSE 損失のみで学習した結果得られる SR 画像との差は見られなかった．

MSE 損失に NIQE 損失を加えて学習して生成された SR 画像については，特に図 7 の SR 画像は，海の波紋や雲の描写を強調するような画像となっている．この画像以外にも，空やほぼ単一色の壁などといったコントラストの変化が乏しい領域のコントラストも大きくするような SR 画像が生成されるということが分かった．ただし，必要以上に画像内のコントラストを大きくしている画像も存在していたため，コントラストの差も明示的に測る SSIM の評価は悪い結果となっている．

MSE 損失に PCA 損失を加えた結果は，他の学習結果と

比べ，輪郭部の先鋭化が他 SR 画像よりも強くされていることが分かる．これにより知覚的精度を大きく向上させることができたのだと考える．ただし，PCA 損失を加えて学習された結果の SR 画像では，色彩の変化が乏しいような領域が，周期的な模様になるような SR 画像となることもあるということも分かった．

5. まとめと今後の課題

本研究では，人の知覚的画質評価を直接的知覚損失関数として実装し，これを用いて人の感覚を機械学習に盛り込み，自動的に知覚的に良い SR 画像を生成するために直接的知覚損失関数の実装を行った．実験では，具体的な効力が明らかになっていない間接的知覚損失関数と直接的知覚損失関数の検証実験を行った．実験の結果，提案した 2 つの直接的知覚損失関数は知覚的精度を向上させることができると分かり，特に PCA 損失は大きく知覚的精度を向上させることができると分かった．しかし，PCA 損失や NIQE 損失の欠点も存在することが分かった．これらの欠点はまだ，効力を把握できていない VGG 損失や GAN 損失，また style 損失を併用することで補えるのではないかと期待される．よって，まずは効力が把握できていない知覚損失関数の影響を把握し，そのうえで，知覚損失関数を複数用いた場合に生成される SR 画像の特徴も把握していく．

また，知覚的精度を大きく向上させることができた PCA 損失の元となった Ma の評価には，PCA 以外に 2 つの処理が存在する．これら 2 つを損失関数化することで SR 画像の知覚的精度を向上させることができることが期待されるため，損失関数として実装し SR 画像の知覚的精度をより向上させていく．

以上のようにして，Pi に基づく知覚的精度をさらに大きく向上させていきたい．その後は Ma らのように被験者実験で実際の人間による主観評価も行うこともしていき，真に知覚的に良い SR 画像を生成していく．

参考文献

[1] M. Irani, and S. Peleg：*Improving resolution by image registration*, Graphical models and image processing(1991).
[2] A. Krizhevsky, I, Sutskever, and G. Hinton: *Imagenet classification with deep convolutional neural networks*, Advances in Neural Information Processing Systems (2012).
[3] C. Dong, C. Loy, K. He, and X. Tang：*Image super-resolution using deep convolutional networks*, IEEE transactions on pattern analysis and machine intelligence(2016).
[4] M. Sajjadi, B. Schölkopf, and M. Hirsch: *Enhancenet: Single image super-resolution through automated texture synthesis*, In:Proceedings of the IEEE International Conference on Computer Vision(2017).
[5] I. Goodfellow, J. Pouget-Abadie,M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, and Y. Bengio: *Generative adversarial nets*, Advances in Neural Information Processing Systems(2014).
[6] C. Ledig, L. Theis, F. Huszár, J. Caballero, A. Cunningham, A. Acosta, A. Aitken, A. Tejani, J. Totz, Z. Wang, and W. Shi: *Photo-Realistic Single Image Super-Resolution Using a Generative Adversarial Network*, In:Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition(2017).
[7] K. Simonyan, and A. Zisserman：*Very Deep Convolutional Networks for Large-Scale Image Recognition*, International Conference on Learning Representations(2015).
[8] A. Mittal, R. Soundararajan, A. Bovik, and C. Bovik：*Making a" Completely Blind" Image Quality Analyzer*, IEEE signal processing letters (2013).
[9] Y. Blau, and T. Michaeli：*The perception-distortion tradeoff*, In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition(2018).
[10] Y. Blau, R. Mechrez, R. Timofte, T. Michaeli, and L. Zelnik-Manor：*The 2018 PIRM challenge on perceptual image super-resolution*, In: Proceedings of the European Conference on Computer Vision(2018).
[11] M. Haris, G. Shakhnarovich,and N. Ukita：*Deep backprojection networks for super-resolution*, In: Proceedings of the IEEE conference on computer vision and pattern recognition(2018).
[12] R. Hassen, Z. Wang, and M. Salama：*No-reference image sharpness assessment based on local phase coherence measurement*, In: 2010 IEEE International Conference on Acoustics, Speech and Signal Processing(2010).
[13] C. Ma, C. Yang, X. Yang, and M. Yang, Ming-Hsuan：*Learning a no-reference quality metric for single-image super-resolution*, Computer Vision and Image Understanding(2017).
[14] O. Russakovsky, J. Deng, H. Su, J. Krause, S. Satheesh, S. Ma, Z. Huang, A. Karpathy, A. Khosla, M.Bernstein, A. Berg, and L. Fei-Fei：*Imagenet large scale visual recognition challenge*, International Journal of Computer Vision(2015).
[15] L. Gatys, A. Ecker, and M. Bethge：*Texture synthesis using convolutional neural networks*, Advances in Neural Information Processing Systems(2015).
[16] L. Gatys, A. Ecker, and M. Bethge：*Image style transfer using convolutional neural networks*, In:Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition(2016).
[17] R. Timofte, E. Agustsson, L. Van Gool, M. Yang, L. Zhang, B. Lim, S. Son, H. Kim, *et al.*：*Ntire 2017 challenge on single image super-resolution: Methods and results*, In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops(2017).
[18] N. Lasmar, Y. Stitou, and Y. Berthoumieu：*Multiscale skewed heavy tailed model for texture analysis*, In: 2009 16th IEEE International Conference on Image Processing (2009).
[19] J. Johnson, A. Alahi, and L. Fei-Fei：*Perceptual losses for real-time style transfer and super-resolution*, In: European conference on computer vision. Springer, Cham(2016).
[20] A. Dosovitskiy, and T. Brox: *Inverting visual representations with convolutional networks*, In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition(2016).
[21] https://github.com/fatheral/matlab_imresize/blob/master/imresize.py.