特集
Special Feature

［知能コンピューティング― AI とハードウェアの出会い―］

3 画像の解像度と知的処理の関係を見つめ直す

―知的な高解像度リアルタイム処理に向けて―

池辺将之　北海道大学 量子集積エレクトロニクス研究センター

画像処理の基本「画像を綺麗に見せること」

スマートフォンや監視カメラに使われているイメージセンサは性能向上が進み，その出力をどのように表現・表示するかも長年研究が続けられてきた．

イメージセンサは，光を電気に変換する画素回路が 2 次元に配列している．光情報として画素回路に届いた光を電子に変換し，アナログ電圧として中間出力した後，アナログ・デジタル変換器によって，画素ごとのデジタル出力値を得る．得られたデジタル画像出力は，信号処理により加工される．このとき，加工した後の画像として人が美しい／綺麗であると感じる処理結果を得ることは，メーカや研究機関に属するセンサ開発や画像処理の研究者らにとって，普遍的な課題として捉えられている．

人が美しい／綺麗であると感じる画像の特徴として主に，画像全体の明暗部のバランスが取れ，細部のコントラストが明瞭であることや不必要な信号成分（ノイズ）が少ないことが挙げられる．そのため，画像の明るさを示す入力輝度を変換し，どのような出力輝度とするかをマッピングする関数の定義付け（トーンマッピング）と画像信号に含まれるノイズを除去する技術は重要である．

一般的に，画像全体の明暗のバランスを調整するにはすべての画素に対して，同じ入出力関数を適用するグローバルな演算，細部コントラストを調整するには局所領域の輝度情報で定義した入出力関数によるローカルな演算がそれぞれ行われる．そのため，この 2 つの演算を適切に扱うことで，画像全体の大まかな特徴を考慮しながら，細かいテクスチャなどのコントラストを明瞭に調整することが可能となる．

また，星空や夜景のように光量の少ない暗い画像はノイズの割合が多く，そこから鮮明な画像を得るにはノイズ除去は欠かせない．イメージセンサ内で発生するノイズについては，物性／回路研究の進展が進み改善されてきた．最終的にノイズ除去の目標として，センサ外の物理的要因によるポアソンノイズが重要課題となっている．

ポアソンノイズは，信号の粒子性すなわち光子数の時間揺らぎに相当するノイズである．例として 1,000,000 個の光子が単位時間で単位面積に到達するような明るい環境下では，その平方根である 1,000 個の光子揺らぎを生じる（ノイズの影響は 0.1%）．100 個の光子が到達するような暗い環境下では，その平方根である 10 個の光子揺らぎを生じてしまう（ノイズの影響が 10% まで上昇）．

トーンマッピングとノイズ除去による「綺麗である画像の取得技術」は，人の思いを美しく残すだけでなく，監視，車載，医療など幅広く使われ，生活

© 2022 Information Processing Society of Japan

を豊かにすることにも繋がっている．最終的に「綺麗」と感じるのは人間であり，人の感性に寄り添う画像の輝度分布／バランスなど，人工知能のような画像の捉え方と処理が重要になってきている．

そのため，これらの技術にも深層畳み込みニューラルネットワーク（DCNN : Deep Convolutional Neural Network）が活用されてきており，画像内のシーンの特徴をも組み入れた処理ができるようになってきた．そして，トーンマッピングとノイズ除去に適したDCNN用の教師画像データも提供されている状態である．ただし，現行DCNNが実時間で処理可能な解像度（例：640画素×480画素）とイメージセンサが出力する解像度（例：4k2k動画3840画素×2160画素）は大きくかけ離れており，画像解像度の扱いと処理負担の関係を精査する必要がある．

本稿では「画像を綺麗に表現すること」の基本的な技術であるトーンマッピングおよびノイズ除去について，処理解像度の関係と深層畳み込みニューラルネットワークを用いた知的処理の今後の展開についても述べていく．

画像変換とCNN

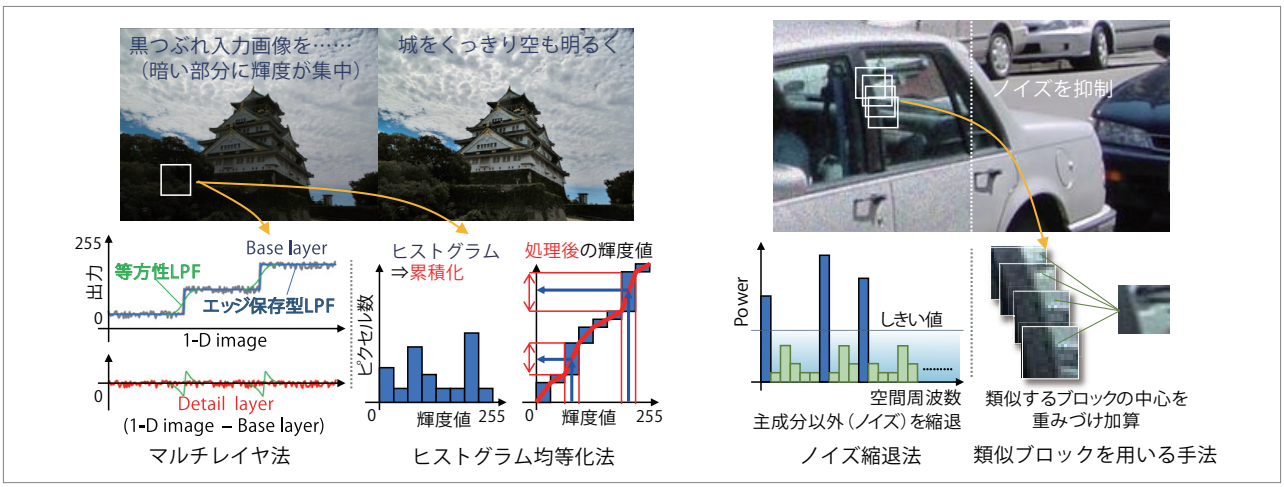
トーンマッピングの理解

外界は輝度レベルとして約10^{-6}カンデラ／m^2（夜空）から約10^6カンデラ／m^2（太陽光）もの広い明暗差（ダイナミックレンジ）があり，その中で20ビット（約10^6）のダイナミックレンジを持つデジタル値として出力できるイメージセンサも開発されてきた．

しかしながら，ディスプレイは，主に10ビット（約10^3）のダイナミックレンジしか扱えない．このとき，外界から取得した輝度情報の範囲を，ディスプレイ表示できる輝度範囲に合わせて，入出力関係を定義し，見た目の良さを保ちながらダイナミックレンジ圧縮する必要がある．

トーンマッピングにおいて，画像全体の大まかな特徴と細かなテクスチャを制御するには，ほぼ大きく2つの手法があり，画像を上記特徴ごとに分割して扱う複数レイヤ法と局所領域ごとの輝度の統計量に基づいた局所ヒストグラム均等化（LHE : Local Histogram Equalization）法が使われている（図-1左下）．

複数レイヤ法では，入力画像に低域通過フィルタ（LPF：Low Pass Filter）を適用し，画像の低周波成分からなるBase Layer，入力画像からBase Layerを差し引いた高周波成分（画像の細かいテクスチャなど）からなるDetail Layerを輝度制御に扱う．上述のダイナミックレンジ圧縮では，室内の窓から晴れた室外を俯瞰する場合，Base Layerの大まかな明暗差（室内と室外の大まかな明るさ）を抑制した上で，強調されたDetail Layer（室内と室外それぞ



■図-1　トーンマッピングとノイズ除去

© 2022 Information Processing Society of Japan

れシーン内のコントラスト）との合成を行う．その結果，本来室内に着目した場合，室外が白飛びしたり，室外に着目した場合，室内が黒つぶれしたりする画像表示を，室内と室外両者ともに適切な明るさで表示できる．

複数レイヤ法は，層分離に単純な低域通過フィルタ（ぼかしフィルタ）を用いた場合，画像中の物体のエッジ付近に不自然さが現れる．これは，ぼかしフィルタが物体境界のエッジもぼかすため，入力画像と Base Layer の差分により得られた Detail Layer にエッジの明暗差情報が含まれてしまうことによる．そのため通常は，エッジ保存型低域通過フィルタを用いて Base Layer を得る．

また，各レイヤをどのくらいの大まかさ細かさで取得するかのフィルタの帯域制御，細かさをどの程度はっきりさせるかの Detail Layer の増幅度，各レイヤの合成比率なども考慮しなければならない．複数レイヤ法を適用するためには，適切なパラメータを自動的に設定する仕組みが必要である．

ヒストグラム均等化法は，画像全体や局所領域の統計量に基づいた非線形トーンマッピングの 1 つで，輝度の情報をヒストグラム化して，処理後のヒストグラム分布が均等になるよう処理が施される（偏った輝度の情報が表示範囲に均等に分配される）．例として図 -1 左上の城の画像の輝度分布について考えてみる．城の部分は黒つぶれしており，輝度が低い領域に特定の輝度が集中している．ヒストグラム均等化は，特定の輝度値に集中してしまった画素群を広い表示範囲に割り当てる．その結果ヒストグラムの分布が広がって均等化する．

ヒストグラムを累積し，縦軸を表示範囲に正規化した関数は，この条件にあてはまり，そのままマッピング関数となる．局所領域ごとにヒストグラム均等化を行う場合，グラデーションなどもディスプレイ表示範囲全体に輝度を割り当てるので過強調を起こしやすく，同時に膨大な計算量が課題となる．過強調はヒストグラムをあらかじめ滑らかにしておくと均等化した後のヒストグラム形状との差が抑制されて改善する（Smoothed LHE）．

ARM 社が買収した Apical Iridix[1)] は，局所ヒストグラム均等化に似た手法を採用しており，スマートフォンや市販カメラの画像補正において市場独占に成功している．Iridix では，フーリエ級数展開のように直交変換された局所ヒストグラムからマッピング関数を再構成する．低周波側から 4 ～ 6 個の直交基底関数で構成された滑らかな累積ヒストグラムを直接計算して，演算時間と過強調の問題を同時に解決している．

ノイズ除去の理解

画像のノイズ除去は，トーンマッピング同様に従来から研究されてきた普遍的な技術課題である．基本的には画像の特徴とノイズ成分を分離して，いかにノイズ成分だけを抑制するかが鍵となる．画像の特徴はテクスチャやエッジ成分に存在するとされ，いままでエッジ保存型低域通過フィルタがこの分野でも使われてきたが，トーンマッピングの Base Layer 抽出のように大まかな物体領域を指定するようなマスキング用途の応用へと移っていった．

そのため，ノイズ抑制に主に使われる技術は直交変換により得られた画像の空間周波数成分から，画像を構成する成分とノイズ成分を分けて，ノイズ成分をシュリンクする技術や画像の中で似ているテクスチャを持つ微小領域を重ね合わせて平均化することで画像の特徴を強調しながらノイズのみを抑制する手法などが用いられている（図 -1 右）．

CNN と出力解像度

畳み込みニューラルネットワークを利用した画像変換というタスクでは，演算負荷と出力解像度の関連性を考える必要がある．画像変換は，カラー化／白黒化，欠損補完，超解像，絵画風など見え方を変更するスタイル変換など，多くの一般的なコンピュータビジョンの応用を実現するための重要な技

© 2022 Information Processing Society of Japan

術である．ここでの課題は，ある画像セットから別の画像セットへのマッピングを学習・推論することであり，画像セットの特性はどのような画像変換を目指すかに依存する．トーンマッピングやノイズ除去の目的は，画像解像度を保ちながら，与えられた画像を変換することであり，物体を認識して分類することではない．

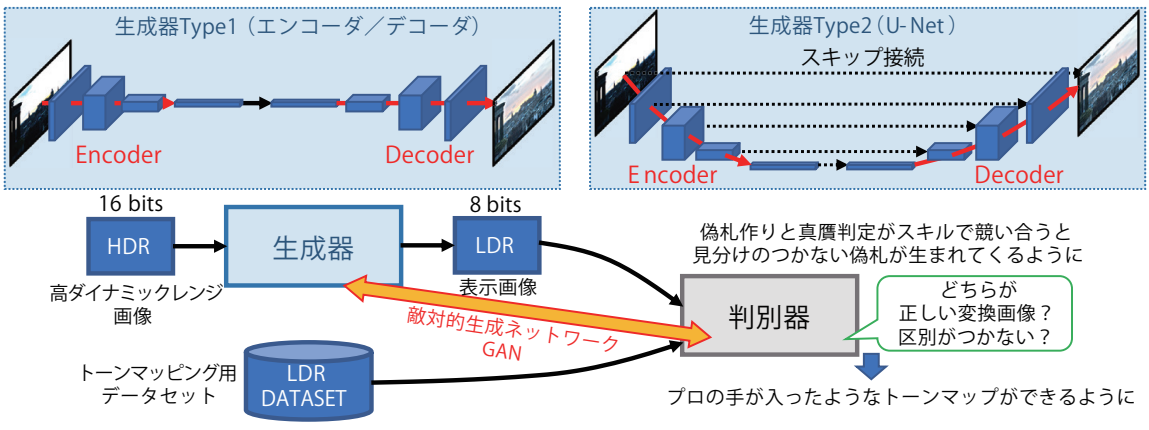
画像変換のための典型的なネットワーク・アーキテクチャは，畳み込みニューラルネットワーク（CNN : Convolutional Neural Network）を基にしたエンコーダ・デコーダである．画像変換では，特徴を得るだけでなく，その特徴を再利用して，画像を再構成しなければならない．

図 -2 左上に示すようにエンコーダは，画像の低レベルの特徴の空間的な寸法を縮小するプーリング畳み込み層を配置して特徴を段階的に統合していく．特徴の統合により高レベルの複雑な特徴を得ることが可能な一連の畳み込み層を実現する．デコーダは，入力―出力（ターゲット）画像の解像度（超解像などのタスクでは出力解像度が異なる場合がある）に合わせて特徴の空間的な寸法を拡大して再現していく．統合された特徴の再現について，拡大する逆畳み込み層を配置した層構造を組み合わせて用いる．

単純な CNN モデルでは，画像の高周波特性が失われ，ぼやけてしまう可能性があり，同一解像度の層間で情報の伝達を行うためのスキップ層を持つ U-Net（図 -2 右上）や敵対的生成ネットワーク GAN（Generative Adversarial Networks）が用いられる．GAN は，画像を生成するネットワークとその生成された画像が正解画像に即して，同じ特徴を有しているかを判別するネットワークからなる．この 2 つのネットワークを競い合わせて学習する．

最終的に判別ネットワークが生成画像と正解画像を区別できなくなったならば，その生成ネットワークは正解画像と同じ特徴を持つ画像を生成し得る能力を持つことになる．紹介したネットワーク構造は，入力と出力同じ解像度の出力画像を再構成するため，認識・分類用ネットワークと比較して構造が大きくなり演算量が増えてしまう（図 -2 下）．

トーンマッピングを CNN で実現するには，エンコーダ・デコーダ構成または U-net を生成ネットワークとして GAN の中に組み込み実装する．判別ネットワークは物体認識・分別用のネットワークを使う．トーンマッピング用 CNN を学習するには，外界の情報に相当する高ダイナミックレンジ画像とそれをプロの写真家が修正してディスプレイ表示可能な輝度範囲に落とした画像をセットにして行う．その結果，生成ネットワークは，高ダイナミックレンジ画像に対して，あたかもプロの写真家が修正したかのような画像変換を行うことができるようになる．



■図 -2　トーンマッピングと深層畳み込みネットワーク

© 2022 Information Processing Society of Japan

トーンマッピングと CNN

CNN との高効率処理の相互変換

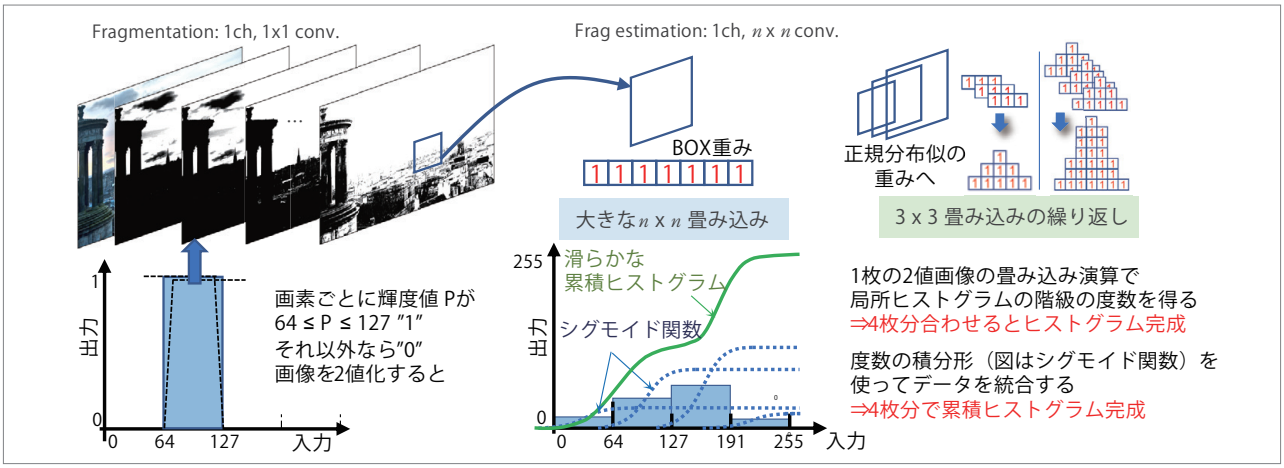
CNN ベースのトーンマッピングを考えた場合，イメージセンサ出力は現在，静止画で数千万画素レベル，動画では 4K2K 解像度が一般的となり，その解像度でのリアルタイム処理（4K 解像度 @60 フレーム／秒）が求められる．現在のハードウェア構成ではこの要求を満たすことはきわめて難しい．そのため，高解像度で高速性を考慮した新しい画像変換処理用のアーキテクチャとアルゴリズムが必要となる．

もし，トーンマッピングの基本要素と CNN の関係性が明らかになれば，高解像度の計算負荷の高い CNN 演算を従来の高解像度向けの高効率処理に相互変換することができる．そして余った演算能力を識別など高度な演算に振り分けることにも繋がる．ここでは，局所領域ごとのヒストグラム均等化（LHE）と CNN の関係性について考えてみよう．LHE の出力は非線形関数である累積ヒストグラムである．輝度階級ごとに得られた度数を積分（累積）していくことと CNN の関係性を考えなければならない．

まず図 -3 の左上に示すように，入力画像を輝度階級ごとに分割してみよう．トーンマッピングでは入力として 8 ビット以上の画像を扱うことが多いが，ここでは 8 ビット（0 ～ 255 までの輝度範囲）とする．そして，簡単のため 4 つの階級（0 ～ 63，64 ～ 127，128 ～ 191，192 ～ 255）を持つヒストグラムの生成を考える．画像全体に対して 64 以上 127 以下に当てはまる画素だけ 1 にしてほかは 0 とするとその条件を反映した 2 値の画像ができあがる．同様に，それぞれの階級の条件に沿った 2 値画像を作ることができる．この作業は，1 画素単位の条件を全画面に適用した処理となっている．

次にここで得られた 4 枚の画像に対して図 -3 上に示すように $n \times n$ の領域で 1 の画素を集めてみよう．これは $n \times n$ の領域で重みが等しい BOX フィルタをかけることと同じである．一方，3 × 3 の BOX フィルタを繰り返し行うと中心極限定理から，5 × 5 の領域サイズで中心に空間重み持つフィルタなり，次に 7 × 7 の領域サイズで正規分布を近似した空間重みを持つようなフィルタとして局所領域内の 1 の数を積算できる．

さらに図 -3 右下に示すように，特定の領域に対して，4 枚の 2 値画像から得られた情報を統合すると，4 階級を持つ局所ヒストグラムが完成する．情報統合の前に，1 の数の積分値（積分関数としてシグモイド関数や Relu6）と階級ごとのオフセットを与えれば，直接，累積ヒストグラムが求まってしまうのである．累積ヒストグラムはトーンマッピング



■図 -3　トーンマッピングと CNN の相互変換

© 2022 Information Processing Society of Japan

のマッピング関数であることは前章で述べたとおりである．このようして，トーンマッピングの基本要素である LHE は画素ごとの演算と畳み込み演算および結果の統合処理に相互変換が可能なのである．

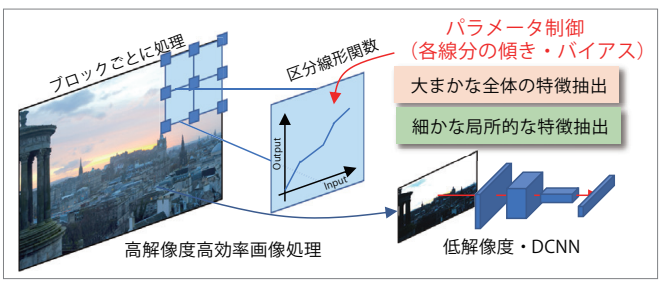
解像度を考慮したアーキテクチャの工夫とリアルタイム処理

CNN を活用したトーンマッピングは，入力画像の解像度を保持する必要があり，高解像度の実時間処理には工夫が必要である．しかし，現行の画像変換用のネットワーク・アーキテクチャでは高解像度による演算負荷で実現が厳しい．

そこで，低解像度 DCNN により画像の特徴を得ることで，高効率な高解像度トーンマッピング処理の制御を行うアーキテクチャを考察してみる．

Gharbi らの HDR-Net ではバイラテラルグリッドと呼ばれる高効率な高解像度トーンマッピング処理手法に対して，低解像度 DCNN によるパラメータ制御を行っている．バイラテラルグリッドでは，ブロックごとに分割された局所領域ごとに，累積ヒストグラムに似たパラメトリックな区分線形関数を用いている．図 -4 に示すように，低解像度 DCNN を用いて画像の大まかな／細かな特徴を抽出して，バイラテラルグリッドにおける区分線形関数の各線分の傾きとバイアスを制御する．

この構成で使われる DCNN は低解像度で扱っているため，演算負荷が小さいとともに出力がパラメータであるためエンコーダも必要としない．DCNN と高解像度処理とのデータ転送量も削減することができ，知的かつ高解像度な実時間トーンマッピングが実現できる．



■図 -4　低解像度 DCNN によるトーンマッピング

ただし，HDR-Net のようなシステムを実現するためには，片翼であるバイラテラルグリッドのような高効率な高解像度処理アルゴリズム，高解像度処理の学習用データセットとアーキテクチャ全体を通した学習方法が必要であることに注意すべきである．

ノイズ除去と CNN

CNN による実装

CNN を活用したノイズ除去は，ノイズ自体の推定を行うものが多い，推定されたノイズと入力画像の差分からノイズのない画像を得る．また，学習のためノイズを含む／含まない画像セットを GAN で作る試みもなされている．また，U-net のダウンサンプリング機構を活用して，細かい特徴から大まかな特徴に対して直交変換を模した解析からノイズと画像の主要成分を分離する．画像の中で似ているテクスチャを持つ微小領域グループを学習に用いて，ノイズ成分を推定する手法も提案されている．

U-net による実装と処理解像度

ここでは，U-net を用いた手法について考えてみよう．まず，低解像度および高解像度の画像にノイズが乗っている状態を想像してほしい．低解像度の画像では，画像の細かい特徴がつぶれて，ノイズとの違いが見つけにくいのではないだろうか？　逆に高解像度の画像では，ノイズが含まれていても元の構造を想像しやすい．

これは，解像度が高いとノイズの拡がりとそこに埋もれている特徴を分離しやすくなり，画像の一部を切り出してもノイズの推定ができることを意味する．すなわち，画像をブロックごとに分けて，それぞれにブロックサイズの小規模ネットワークで処理できることを示唆しているのである．そして，ブロック処理用のネットワークはサイズが小規模なので，

© 2022 Information Processing Society of Japan

複数の処理ユニットを用意することや処理ステージを分けてステージ間に一時的な記憶領域を設けることでパイプライン化も可能となる．これは小規模ネットワークだから実現可能な構成である（図 -5）．

処理画像と見た目

ノイズ除去の性能向上のため，U-Net を複数個，カスケード接続してみよう．このとき，初段 U-Net で得られた出力に含まれるノイズの特徴は，入力画像のノイズが持っていた特徴と異なる．そのため，次段 U-Net は新しいノイズ成分に対して学習し，その除去を行うこととなる．ただし，カスケード接続したネットにより出力された画像は，ノイズが減って滑らかな画像に作り替えられる．これは，入力画像が持つ画像の特徴さえも整形して見た目を良くしてしまう可能性があるのだ．そのため，ノイズ除去と画像の見た目の判断に新たな指標が必要となる．次節で，見た目にかかわる人の感性と CNN の関係について考えてみよう．

人の感性と CNN

人の感性によるデータセット構築

トーンマッピングやノイズ除去は，主観である見た目と大きくかかわっている．トーンマッピングでは，画像の自然さを損なわずに見た目を向上させることが求められる．ノイズ除去も同様である．

そのため，データセットの作成において，Bychkovsky らは一眼レフカメラを用いて，さまざまな風景，被写体，異なる照明条件で 5,000 枚の写真を撮影し，それらの写真を，5 人のプロの写真家の協力を得て，Adobe Lightroom を使って手動により色調を調整したデータセットを作成している．

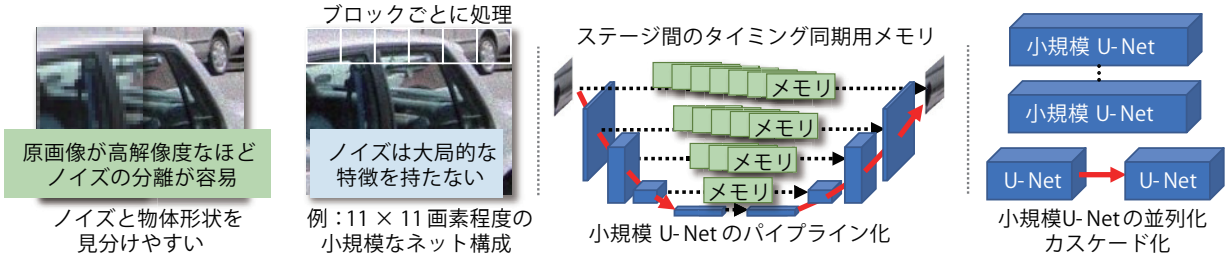
最近では，Panetta らが，456 枚の高ダイナミックレンジ画像からなるデータセットを公開している．これらの画像は，屋内，自然，夜間，川辺の夕日の 4 つのカテゴリに分類されており，さまざまな照明条件で撮影され，高ダイナミックレンジ画像として，Photoshop の機能を用いて，5 つの異なる露出を融合して作成されている [2]．

感性の定量化と今後の展開

自然さや見た目は主観的要素が強く，これらの定義や定量評価は難しい．医療画像では見た目の良い画像変換を行うことで，画像診断 DCNN の性能向上にもつながる報告もなされている．この場合，トーンマッピング機能やノイズ除去が DCNN の一部として組み込まれる．

「自然さ」という概念を深堀りするには，第一にガラスや金属などの個々の物体の材質がなぜそれっぽく見えるのか，画像内のマテリアルの特徴量の問題として考えることができる．そして，マテリアル特徴解析の分野は，議論が活発になってきている（Motoyoshi ら 2007 [3], Sawayama ら 2017 [4]）．

これらの研究は画像の印象から「金属っぽさ」などの特徴量をもとに材質変換にも研究を進めている．筆者らはその知見を活用し，いかに画像内のマテリアル群の特徴量の統合的に定義するかと，いかに統



■図 -5　小規模 DCNN（U-Net）を用いたノイズ除去用アーキテクチャ

© 2022 Information Processing Society of Japan

合的な特徴を保ったまま画像変換できるかに着目しており，質感を包含する「自然さ」という，より大きな概念を数理記述することを目論んでいる．

そして，画像の見た目については錯視も関連している．例として皆さんが一度は目にしたことがあるチェッカーシャドウ錯視など，物体の模様と陰影などの状況から引き起こされる錯視はまさに人の内部で生み出された感じ方と言えよう．DCNN と錯覚においては，蛇の図柄がなぜ回転するように見えるか[5]，その特徴量の抽出に成功している．同時にさまざまな錯視を GAN で生成しようという試みもなされているが．こちらは失敗しているようである．

さて，人の感じ方が大きくかかわる錯視の効果は定量化できるのであろうか？　ここでは，文系の学術分野である心理物理学がその定量化のノウハウを有している．金子らは，実験心理学的研究により，「見え」の主観を定量化しヒトの視覚の情報処理メカニズム（特に低次視覚特徴の処理機構）の解明を目指してきた．特に錯覚の発生機序に関する研究を多数行っている[6]．

自然さや見た目など，人の感性がかかわる主観的要素の解析は，文工の学術分野における新しい連携の形を示す．それとともに解析の結果は，人が考えるより大きな概念の定量化に結び付いて，それを利用した AI との新しい関係や人の感性を模したデータセットの生成，人に近い画像変換システムの最適化につながっていくと考えている．幅広い学術分野の連携と人を主体とした新しい AI 産業の発展を強く望むとともに，その研究を進めていきたいと考えている．

参考文献

1) Chesnokov, V. : Image Enhancement Methods and Apparatus Therefor, US Patent 7,302,110 (2007).
2) Panetta, K., Kezebou, L., Oludare, V., Agaian, S. and Xia, Z. : TMO-Net : A Parameter-Free Tone Mapping Operator Using Generative Adversarial Network, and Performance Benchmarking on Large Scale HDR Dataset, IEEE Access, 9, 39500-39517 (2021).
3) Motoyoshi, I., Nishida, S. Y., Sharan, L. and Adelson, E. H. : Image Statistics and The Perception of Surface Qualities, Nature, 447(7141), 206-209 (2007).
4) Sawayama, M., Nishida, S. Y. and Shinya, M. : Human Perception of Subresolution Fineness of Dense Textures Based on Image Intensity Statistics, Journal of vision, 17(4), 8-8 (2017).
5) Watanabe, E., Kitaoka, A., Sakamoto, K., Yasugi, M. and Tanaka, K. : Illusory Motion Reproduced by Deep Neural Networks Trained for Prediction, Frontiers in psychology, 9, 345 (2018).
6) Kaneko, S. and Gilchrist, A. : Lightness in a Flash : Effect of Exposure Time on Lightness Perception, i-Perception, 11(6), 2041669520983830 (2020).

（2021 年 12 月 6 日受付）

■池辺将之　ikebe@ist.hokudai.ac.jp

2000 年，北海道大学大学院電子情報工学専攻博士課程修了．2000 年，大日本印刷（株）半導体製品研究所．2004 年，北海道大学大学院情報科学研究科准教授．2018 年，北海道大学量子集積エレクトロニクス研究センター教授．2021 年，北海道大学病院次世代遠隔医療システム開発センター副センター長兼任．

© 2022 Information Processing Society of Japan