多露光画像の合成における深層学習を用いたアーティファクト抑制に関する研究

船橋 勇那

電気通信大学大学院 情報理工学研究科
情報・ネットワーク工学専攻
博士（工学）学位申請論文

2022 年 3 月

多露光画像の合成における深層学習を用いた
アーティファクト抑制に関する研究

博士論文審査委員会

主査　吉田太一　助教
委員　張　熙　教授
委員　野村英之　教授
委員　高橋弘太　准教授
委員　劉　志　准教授

Abstract

Since image are taken by ordinary digital cameras with low dynamic range（LDR），they have over- and under-exposure regions for high dynamic range（HDR）scene. Over- and under-exposure regions have saturated pixel values，and lose information of image features that are edges or detail features of objects. The loss of information causes problems on technics of image application，for example image recognition，object recognition，depth estimations from images，and so on. To solve the problems，HDR image fusion and multi-exposure image fusion technics have studied. These technics fuse the multi-exposure images that are taken by various exposure settings，and provide a HDR image or an image without saturated values of pixels. However，artifacts occurs in an image produced by fusing a set of multi-exposure images in which locations of objects and details are different. For avoiding the artifacts，I propose image adjustment method of the multi-exposure images based on deep learning in this study. The multi-exposure images that taken in natural scenes has regions are lost the information of image features due to combination of the saturation of pixels and the object moving. Since the regions causes artifacts of the resultant image，image adjustment method with considering the regions can avoids the artifacts. The proposed method set the reference image from the input multi-exposure images，and adjusts objects and details in multi-exposure images to the reference one. The proposed method detects and inpaint the regions based on two convolutional neural networks. As a result，the proposed method provides adjusted multi-exposure images that avoids the artifacts. In the experiments of the quantitative evaluation and the visual evaluation，it is observed that a simple fusion method with the proposed method outperforms state-of-the-art fusion methods，and the proposed method show obviously reducing artifacts.

要旨

一般に普及しているデジタルカメラのイメージセンサは記録できる輝度のダイナミックレンジが低く，撮影された画像には白とびや黒つぶれといった欠損領域が発生する．そのような白とびや黒つぶれの欠損領域は，画像の画素値が飽和することで物体の画像特徴となるエッジや模様といった情報が失われてしまっている．そのため，カメラをセンサとして撮影画像を応用する画像認識や物体認識技術，深度推定などの応用技術において悪影響を及ぼす．その問題を解決するため，複数枚の露光の異なる画像を合成する多露光画像合成技術および High Dynamic Range（HDR）画像合成技術が研究されてきた．それらの技術は多露光画像という複数枚の露光を変えながら撮影した画像を合成することでダイナミックレンジが高い画像を合成できる．しかし，多露光画像合成や HDR 画像合成では，複数枚の画像を合成するため物体が二重に現れるゴーストと呼ばれるアーティファクトが発生する．

本研究では，多露光画像合成および HDR 画像合成のアーティファクト抑制を目的として，深層学習を用いた画像内物体の位置補正を適応的に行う画像補正手法を提案する．従来のアーティファクト抑制手法では，多露光画像それぞれの白とびおよび黒つぶれと画像間の物体の位置ずれが同時に発生している場合にアーティファクトが発生してしまっていることがわかった．それらの領域を定義すると，基準画像において画素値の飽和である白とびが発生し基準画像よりも露光が低い画像では画像間の位置ずれが発生している領域であり，逆に基準画像において黒つぶれかつ基準画像よりも高い露光では画像の位置ずれが発生している領域である．それらの領域では，多露光画像において画像情報が失われている．ここで白とび黒つぶれとは画像の RGB すべての値が飽和している領域である．そのため，従来法で用いられているオプティカルフローやパッチ画像のマッチングによる位置ずれ計測では計測の元となる画像情報がなく位置ずれ計測が失敗し結果としてアーティファクトが発生する．

そこで，本研究では従来法の位置ずれ補正に加え，定義した欠損領域を適応的に検出し，画像情報を補間する新しい手法を提案する．従来法では位置ずれの補正と画像合成時にアーティファクト抑制を行っている．しかし，アーティファクトは画像間で物体の位置ずれと前述した欠損領域の補間ができれば発生せず，合成時の抑制処理を必要としない．よって，本論文では多露光画像間の位置ずれを補正する深層学習を用いた手法を提案する．

提案手法は，多露光画像から 1 枚を基準画像とし他の画像内の物体の位置ずれの補正を行うことで位置ずれのない多露光画像を生成する．定義した領域は，基準画像ともう 1 枚の画像を基準画像とその他の露光の画像 2 つを用いて検出できる．よって提案法は，基準画像とそれ以外の露光の 2 枚を入力として補正し，基準画像以外のすべての露光の画像に補正を行うことで位置ずれを補正し補間した多露光画像を作成する．具体的には，従来法を基に 2 枚の画像間で位置ずれを補正し，次に定義した領域の検出を提案する検出用 Convolutional Neural Network（CNN）モデルで検出し，さらに提案する画像補間

CNN モデルにより補間する．それらにより，画像間で位置ずれがなくアーティファクトを発生させる原因となる領域を削減した多露光画像を作成する．提案法で作成された位置ずれなしの多露光画像を画像合成に用いることにより，結果としてアーティファクトを抑制した合成画像が得られる．また，公開されたデータセットを用いて提案手法の評価実験を行った．従来のアーティファクト抑制処理を含む画像合成手法の結果と定量的および視覚的に評価する比較する実験を行った．比較実験より，提案手法を合成手法の前処理として適用することにより従来法に比べ高いアーティファクト抑制効果を得られることがわかった．

目次

図目次

表目次

第 1 章

序論

1.1　本研究の背景

近年，デジタルカメラやその処理のための高性能な演算用大規模集積回路などのハードウェアと保存や加工を行うソフトウェアが盛んに研究開発され実用化されており，誰もが少なくとも 1 台は持っていると言えるほどに，視覚的に実環境の情報をデジタル信号として取得するデバイス，つまりデジタルカメラやスマートフォンなどは普及している．また，それらはただポートレート写真を撮影するだけではなく，画像情報を用いた製品検査や警備，ロボットの制御，セルフレジスター，スマートフォンを用いた電子決済などに用いられ，それらデジタル情報処理は私たちの身の回りの生活で広く使われ社会に必要不可欠な技術となっている．

現実世界の情報を取得し保存するセンサ性能として重要なのは情報を正確に記録することである．デジタルカメラは他のレーザー距離センサなどのセンサに比べ，撮影物体との距離や撮影範囲などを柔軟に設定でき広い範囲の密な情報を一度に取得できる点で優れている．デジタルカメラで取得できる情報が高精度かつ正確であれば，画像情報を用いた自動ロボットやコンピュータビジョン技術において所望の動作を正確に実現できる．またカメラで取得できる情報を増やすことで，これまで使えなかった新たな環境や目的で用いられるようになるなど更なる技術の発展や便利な社会の実現に寄与することが期待される．

しかし，現在一般的に普及しているデジタルカメラはその性能に制限があり取得した画像情報に欠損が発生する．デジタルカメラは，現実世界の光の強度である輝度やその波長の違いとして現れる色のアナログな情報をデジタル化して保存するため，カメラの性能により上限と下限を持った離散的な値になる．特に，一般的に普及しているデジタルカメラのイメージセンサでは一度の撮影で取得できる光の強度の幅であるダイナミックレンジが現実シーンや人の視覚特性におけるダイナミックレンジよりも低い．それが制限となり現実シーンのダイナミックレンジをすべて一度に取得することができない．結果として得られた画像は現実シーンの光の強度の一部を切り取って取得しており，上限以上と下限以下の強度の光はすべて画素値の上限値と下限値に丸め込まれて一定値と

して記録される．それら一定値の画像領域は，上限値で記録された場合は白くとび下限値で記録された場合は黒くつぶれており，現実シーンの輝度情報を正確に取得できていない情報欠損が発生した領域となる．

そこで，複数枚の画像を撮影しそれらをデジタル信号処理により合成して，カメラ性能以上の高いダイナミックレンジを有する画像を実現する技術が研究されてきた．その複数枚画像は，現実の同一シーンを異なる多数の露光で撮影することで取得され，多数の露光で撮影された画像群であることから多露光画像と一般的に呼ばれる．多露光画像は，それぞれの画像は低いダイナミックレンジで撮影されたものであり，異なる露光で取得されていることからそれぞれ異なる場所に白とびや黒つぶれが発生する．よって，それら画像間で情報を補い合うことで高いダイナミックレンジを記録できる．その多露光画像の応用技術としては High Dynamic Range（HDR）画像合成や多露光画像合成などがあげられる [3–21].

HDR 画像合成や多露光画像合成は，多露光画像を用いて高いダイナミックレンジを有する 1 枚の画像を実現する技術である [3–21]．それら技術は，非常に簡略化すると多露光画像を画素単位で重みつき合成により合成し 1 枚の画像を作成している．結果として白とびや黒つぶれがなく通常のカメラでは取得できない高いダイナミックレンジを持つ画像を作成できる．しかし，合成処理により現実にはない視覚的劣化（アーティファクト）を人工的に作成してしまう問題や多露光画像のノイズが増幅してしまう問題がある．

それらの問題のうちアーティファクトは現実世界にないものが画像に発生しているため応用において特に影響の大きい問題である．一般的に多露光画像は，連写によって撮影されかつ露光時間が異なるので各画像の取得タイミングが異なり，撮影枚数が多ければ多いほど 1 枚目の画像と最後に取得された画像では撮影された時間のずれが大きくなる．そのため複数枚画像間で，撮影物体自身に動きが発生したり手ぶれなどによって撮影物体の位置が微妙に異なったりする．それら位置ずれが発生した場合，主に画素ごとに合成処理を行うため HDR 画像合成や多露光画像合成では位置ずれした物体が半透明や 2 重になるといった合成処理によりアーティファクトが発生する．アーティファクトが発生した画像は現実シーンの情報を正確に記録しているとはいえないだけでなく，画像認識や画像を用いた深度推定などで誤検出や未検出を引き起こす原因となる．

そこで近年，アーティファクト抑制処理を含む多露光画像の合成手法と多露光画像の補正手法が提案されている [3, 6, 10, 11, 13, 14, 16–24]．従来の手法では，パッチ画像を用いたマッチングにより物体の位置ずれを考慮して画像を合成する手法が提案されている [6, 10]．それらの手法では，パッチ画像と呼ばれる小さな画像を各露光の画像において白とび黒つぶれによる欠損のない領域から選びそれらを用いて合成し欠損領域のない画像を作成する．最新手法では，畳み込みニューラルネットワーク（Convolutional Neural Network: CNN）を用いた手法が提案されている [17–21]．それら最新手法では，多露光画像の補正や合成，重み付き合成の重み推定に CNN を用いており，従来の手法に比べ基本的には良い結果を示している．しかし，それらの手法においても基準画像において白とび黒つぶれが発生し，その他の露光の画像においても同じ領域に白とび黒つぶれおよ

び物体の位置ずれが発生した多露光画像の画像間においても情報が欠損した領域においてアーティファクトが発生する場合がある．特に，多露光画像から学習済み CNN モデルにより直接 HDR 画像や多露光画像を推定する場合，入力できる多露光画像の枚数が限られるなどの欠点も存在する．よって，多露光画像合成および HDR 画像合成のアーティファクト問題は未だ解決しておらず，より多くの応用技術において多露光画像が応用されるにはその抑制が必要であるといえる．

1.2　本研究の目的

デジタルカメラで取得できる輝度のダイナミックレンジを広げ更なる高性能化を行いこれまで以上に様々な環境下で使えるように発展させるためには，多露光画像の合成時に発生するアーティファクトの抑制が必要となる．しかし抑制処理を含む最新の合成手法 [17–21, 24] では，アーティファクトをある程度抑制することはできているが，特定の領域でアーティファクトを発生させている．その領域とは，基準画像の白とび黒つぶれとそれ以外の露光の画像における位置ずれが同じ領域に発生することで多露光画像の画像間においてその画像情報が欠損した領域である．アーティファクトの抑制問題の解決は，HDR 画像合成や多露光画像合成の技術で視覚的に劣化のない画像を得るためだけでなく，それらの技術で取得した画像を画像認識や画像を用いた深度推定など技術に応用した場合に問題になる．そこで本研究では，その欠損領域によるアーティファクト問題の解決を目的とした多露光画像の補正手法を提案する．本研究では，抑制処理を含む最新の合成手法 [17–21, 24] で考慮されていないアーティファクトの発生原因となっている領域を定義し，その領域を考慮しつつ HDR 画像合成および多露光画像合成技術のどちらにも応用可能な新たなアーティファクト抑制手法を提案する．本研究の抑制手法の提案により，多露光画像を用いる技術の発展だけでなく前述した画像の応用技術の発展に寄与できる点で意味のある研究であると考えている．

1.3　本研究の新規性および成果

本研究の新規性は，従来法で問題となっている多露光画像の欠損領域を定義した点と、深層学習を用いた提案 CNN モデルによる多露光画像の欠損領域検出と補間による補正手法を提案した点である．問題となる領域の定義および議論は 4.2 節において行っている．近年，深層学習を用いてアーティファクトを抑制した HDR 画像や多露光画像合成の画像を推定する手法が提案されている [17–21]．その中でも本研究では，前述したアーティファクトが発生する欠損領域を CNN モデルを用いて検出し補間する新たな多露光画像補正手法を提案する．本研究の成果は，次の点である．

- 従来法で十分に考慮されておらずアーティファクトの原因となっていた領域を明らかにした．

- 欠損領域の検出 CNN モデルと提案補間 CNN モデルの 2 つのモデルを組み合わせ適応的に欠損領域を補間する多露光画像の新たな補正手法を提案した．
- 2 枚の画像を入力とする多露光画像補間のための CNN モデルを提案した．
- 検出 CNN モデルと提案補間 CNN モデルの 2 つのモデルを同時に用いて学習させ，検出 CNN モデルのための教師データなしに CNN モデルのパラメータを学習させる学習法を提案した．

1.4　本論文の構成

本論文の構成について記す．第 2 章は，本論文全体で用いる基礎理論を示す．主にデジタルカメラの特性，提案法に用いる技術について述べる．第 3 章は，アーティファクト抑制とそれに関連した従来研究とその手法について述べ，本研究の位置づけを示す．第 4 章は，本研究で提案する深層学習技術を応用した新たなアーティファクト抑制手法について述べる．第 5 章は，提案法と従来法のアーティファクト抑制効果において定量的および視覚的な比較実験を行い，各結果に対して考察を述べる．最後に第 6 章において本研究の結論を述べ，本論文を結ぶ．

第 2 章

基礎理論

2.1 概要

本章では，本論文を理解するために必要な，デジタルカメラや HDR 画像，機械学習と深層学習に関する様々な基礎理論について概説する．2.2 節では，一般的に普及しているデジタルカメラの特性について述べる．2.3 節では多露光画像および HDR 画像について述べた後，2.4 節において HDR 画像を取得するための技術と多露光画像合成技術について述べる．2.5 節では機械学習の基礎理論を述べ，2.6 節では提案法で用いる機械学習技術の 1 つである深層学習について基礎理論を述べる．特に，2.6.1 項では本論文で主に用いている深層学習技術の畳み込みニューラルネットワーク (CNN）について述べ，続いて CNN で用いられるニューラルネットワークの構造や畳み込み演算手法について説明する．2.7 節では，2 枚の画像間での画素ごとの移動量を推定するオプティカルフロー推定について述べる．最後に，2.8 節では画像処理における出力画像の画質評価指標として一般的に用いられる定量評価指標について述べる．

2.2 デジタルカメラの特性

一般的に，市販されている一眼レフデジタルカメラなどの撮像デバイスで取得できる輝度のダイナミックレンジは実際のシーンのそれよりもはるかに低い．実際には光の強さは，例えば曇天夜の屋内では 10^{-5} cd/m^2 程度であり，太陽光下では 10^{9} cd/m^2 程度と非常に高く，そのダイナミックレンジは 200 dB を超える高さを持つ [25]．ここでのダイナミックレンジは下記の式で表される [25].

$$DynamicRange = 20 \log_{10} \frac{I_{\max}}{I_{\min}}, \tag{1}$$

この式で，$I_{\max}$ と $I_{\min}$ は最大と最小輝度値を表す．例えば，照明を点灯させていない暗い室内から窓の外の明るい場所を見るようなシーンは高いダイナミックレンジを持つ場合が多い．近年，イメージセンサの技術研究開発が進み，人の視覚特性とほぼ同等の高いダイナミックレンジを取得できるイメージセンサが開発されている [26]．しかし，一般

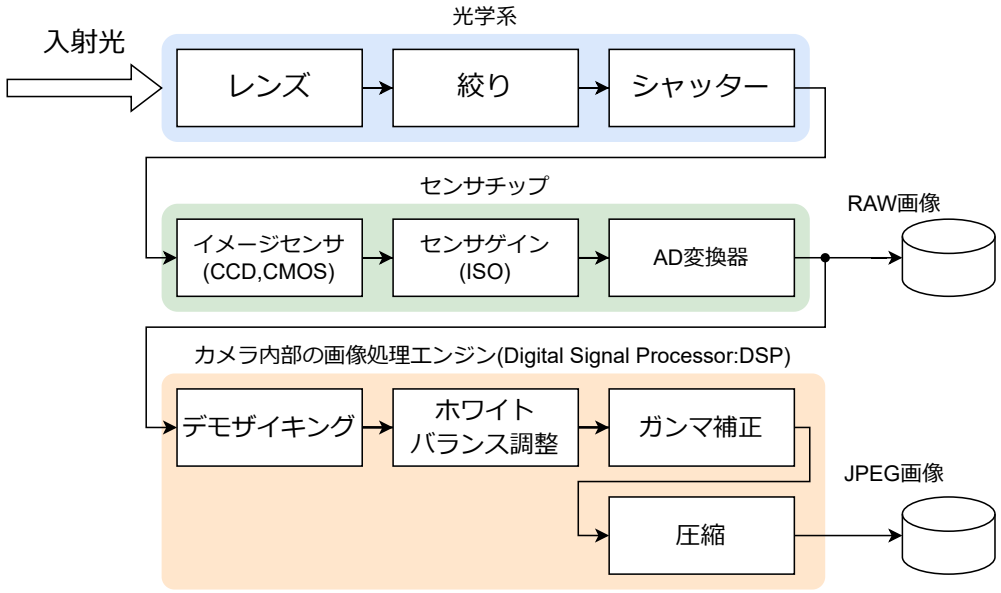


図 1　デジタルカメラの画像撮影時の内部処理

的なデジタルカメラに内蔵されたイメージセンサのダイナミックレンジは 80dB 程度であり，一度の撮影で現実シーンのダイナミックレンジを全て保持した画像を得ることはできない．

一般的なデジタルカメラにおいて，レンズに入力された光から画像情報に変換するまでには複数の処理が適用され，これらの処理はカメラレスポンス関数と呼ばれる非線形関数で近似的にモデル化される．図 1 は，デジタルカメラ撮影時のカメラ内部の処理を表した図である [27]．図 1 に示すように，カメラのレンズから入射した光はレンズや絞り，シャッターなどの光学系機構を通したあと，イメージセンサとその内部の Analog-Digital（AD）変換器により光の強度が量子化される．その後，カメラ内部の画像処理プロセッサによりカラー画像化を行うデモザイキング処理やホワイトバランスの調整，ガンマ補正などの処理が行われ，JPEG などの画像符号化手法で圧縮し保存される．元の光の強度を計算する場合，非線形なセンサ特性やこれらの処理の全てを非線形関数のカメラレスポンス関数として近似的にモデル化する．カメラレスポンス関数は，典型的な近似関数提案されているが，厳密には各カメラによって異なる関数であり関数の形によって色味などが異なる．

低いダイナミックレンジを持つイメージセンサでダイナミックレンジの高いシーンを撮影した場合，白とびや黒つぶれといった画像情報が欠損した領域の発生が問題になる．一般的なカメラで撮影された画像は，通常の JPEG 形式などであれば 0 から 255 の整数 8bit で，RAW 形式のデータであれば 12bit で各 RGB の画素値を記録している．カメラ取得のダイナミックレンジが低いせいで，明るすぎる領域は画素値が 255 もしくは 4095 付近の値となり白くとび，一方で暗い領域は 0 付近の値となり黒くつぶれてしまう．それら画素値が上界および下界で飽和した領域は，現実シーンの情報を正確に記録できてお

図 2　多露光画像

らず画像情報の欠損が発生している．一般的に，画像の応用技術である物体認識や物体検出は，物体の輪郭や模様などの画像特徴を用いて検出する．しかし，白とびや黒つぶれによる欠損が発生した画像の領域では，それらの画像特徴の欠損により誤検出や未検出という問題を引き起こす原因となる．2.4 節で説明する多露光画像合成や High Dynamic Range（HDR）画像合成は，その情報欠損がない高いダイナミックレンジを持つ画像を得る画像処理技術である．

2.3　多露光画像および High Dynamic Range（HDR）画像

多露光画像は，図 2 に示す様な露光設定を変えながら同一シーンを撮影した複数枚の画像群のことである．図 2 は 3 枚の画像群であるが枚数に規定はない．多露光画像は，文献によっては多重露光画像とも呼ばれることもあるが，本論文ではフィルムカメラ等にある 1 枚画像に対して多重に露光を行なう撮影方法で取得された画像と区別するため多露光画像と呼ぶ．多露光画像を取得するには，デジタルカメラのブラケット撮影機能を用いて露光設定を変更しながら連写して撮影するのが一般的である．連写により取得するため，多露光画像の 1 枚 1 枚の画像はそれぞれ異なる時間に撮影され画像間に時間のずれが発生する．その撮影間隔は，設定したシャッタースピードやカメラの内部処理速度により左右されるため一定とは限らない．画像間で撮影時間が少しずつ異なるため，生物や風に揺れる木々などを撮影した場合や手持ち撮影をした場合などに，撮影した多露光画像の画像群において同一の物体が異なる位置に記録されることがしばしば発生する．

HDR 画像は，通常のカメラで撮影されるダイナミックレンジの低い（Low Dynamic Range: LDR）画像よりも高いダイナミックレンジを持ち，現実シーンの輝度情報を保持することを目的とした画像のことである [28]．図 3 に，HDR 画像を適切に LDR 画像のビット幅に変換した画像と同一シーンを撮影した LDR 画像を示す．HDR 画像では，LDR 画像で白とびや黒つぶれが発生している明暗の差が激しい窓の外や人の顔などが，情報欠損なく取得できていることがわかる．HDR 画像は，RGB 各色 32bit 以上の輝度ダイナミックレンジを持つとされている [29]．JPEG や PNG といった通常の画像圧縮フォーマットでは高いダイナミックレンジを保存できないことから，HDR 画像は OpenEXR や RGBE Encoding のようなフォーマットで圧縮し保存される [28]．HDR 画像の取得方法は主に 2 つあり，前述した多露光画像から HDR 画像を合成する手法と，高

図 3　HDR 画像（トーンマッピング後）と LDR 画像

いダイナミックレンジを持つイメージセンサを搭載したカメラでの撮影で得られる．合成により取得する方法は，通常のカメラがあれば多露光画像が撮影可能であるため，非常に高いダイナミックレンジを持つ画像も容易に取得可能である．HDR 画像は，カメラレスポンス関数などが適用された LDR 画像とは異なり輝度に対して線形な値を持つ情報として保存される．HDR 画像を HDR に対応していないディスプレイで映す場合には，人の視覚特性やディスプレイの特性を考慮してダイナミックレンジを適切に圧縮するトーンマッピング処理 [28] を行うことで表示できる．

2.4　HDR 画像合成および多露光画像合成

HDR 画像合成は，多露光画像を用いて HDR 画像を生成する技術である [4, 6, 10, 11, 16–19, 25, 28]．図 4 に HDR 画像合成手法のベースラインとなる手法を示す [4]．2.2 節で述べたように，光がレンズに入射するところから画像とし符号化されるまでの非線形な処理をカメラレスポンス関数としてモデル化される．その逆関数を推定し，画像から現実シーンの輝度分布に近い線形な放射輝度画像を求める．Debevec らの手法では，SVD 法を用いた最適化により逆カメラレスポンス関数の推定を行っている [4]．次に，推定画素値から重み付き合成などにより多露光画像を合成し HDR 画像を作成する．HDR 画像を表示する場合は，LDR 用のディスプレイに適合させるためダイナミックレンジを圧縮するトーンマッピング処理を適用して表示する．

多露光画像合成は，HDR 画像合成と同様に多露光画像を入力として合成処理を行うことで白とびや黒つぶれといった欠損領域のない画像を得る技術である [8, 12]．多露光画像合成は，HDR 画像合成と異なり現実シーンの輝度を表現した HDR 画像の復元を行わずに画素値飽和による欠損領域のない画像を得る．そのため，合成時にカメラレスポンスカーブの推定や表示時の HDR 画像のトーンマッピング処理などが必要ない．よって，HDR 画像合成と比べて処理が少なくてよく，合成後の画像も 8bit の LDR 画像でありロボットのためのコンピュータービジョンなどに応用しやすいという特徴がある．

多露光画像合成のベースライン手法である Mertens らの手法 [8] について述べる．多

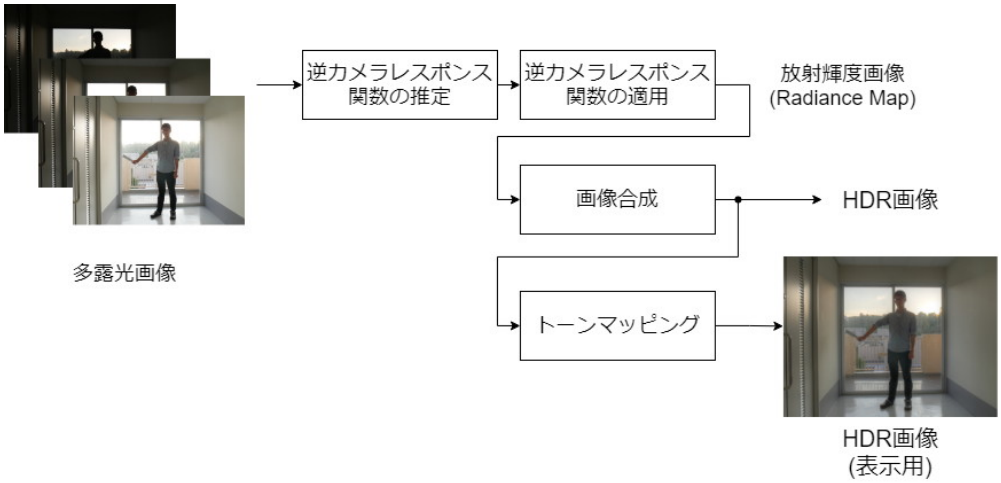


図 4　HDR 画像合成手法

露光画像合成は HDR 画像合成とは異なり多露光画像の各画像から適正露光で撮影された白とび黒つぶれがない領域を選び，合成する．そのため，各画像においてそれらの領域を選ぶために画像品質を測定しそれに基づいて合成が行われる．Mertens らの手法では，入力した多露光画像から各画像のコントラスト，画素値の飽和，Well-exposedness の 3 つの指標から計算した重み合成に用いる．画像合成には算出した画素ごとの重みを用いた重み付き合成が用いられる．ここでは，入力画像の画素値は $[0, 1]$ に正規化されているとする．N 枚の多露光画像を重み付き合成する処理は次式で表される．

$$R_{ij} = \sum_{k=1}^{N} \hat{W}_{ij,k} I_{ij,k}, \tag{2}$$

ここで，i と j は各画素の座標を表しており R_{ij} は合成後の (i, j) 座標における画素値を表している．また，$\hat{W}_{ij,k}$ は k 枚目の多露光画像に適用される重み，$I_{ij,k}$ は k 枚目の画像の画素値を表している．Mertens らの手法では，この画素ごとの重み付き合成をラプラシアンピラミッドと呼ばれる複数解像度の画像群を用いて行う．

Mertens らの手法で算出される重み付き合成のための重みの算出方法について述べる．重みはコントラスト C と画素値の飽和 S および Well-exposedness E の 3 つの項で構成される．まず C は，入力多露光画像の各画像を RGB 画像からグレースケール画像へと変換し，その画像に 3 のフィルタを持つラプラシアンフィルタを適用し絶対値をとった値である．コントラストが高い領域は画像のエッジやテクスチャといった画像の細かい特徴が多く，ラプラシアンフィルタの結果を用いることでそれらの特徴が多い領域の重みの値が大きくなる．次に S は，各画素における RGB 値の標準偏差を用いる．C は次

式で算出される．

$$\mu_{\mathrm{RGB}} = \frac{(I_{ij,k,\mathrm{R}} + I_{ij,k,\mathrm{G}} + I_{ij,k,\mathrm{B}})}{3}, \tag{3}$$

$$C_{ij,k} = \sqrt{\frac{1}{3}((I_{ij,k,\mathrm{R}} - \mu_{\mathrm{RGB}})^2 + (I_{ij,k,\mathrm{G}} - \mu_{\mathrm{RGB}})^2 + (I_{ij,k,\mathrm{B}} - \mu_{\mathrm{RGB}})^2)}, \tag{4}$$

ここで，$I_{ij,k,\mathrm{R}}$，$I_{ij,k,\mathrm{G}}$，$I_{ij,k,\mathrm{B}}$ は各画素の RGB 値をそれぞれ表している．最後に Well-exposedness E は RGB の画素値と 0.5 との近さを次式で算出し，その値を用いる．

$$E_{ij,k,\mathrm{R}} = \exp -\frac{(I_{ij,k,\mathrm{R}} - 0.5)^2}{2\sigma^2}, \tag{5}$$

$$E_{ij,k,\mathrm{G}} = \exp -\frac{(I_{ij,k,\mathrm{G}} - 0.5)^2}{2\sigma^2}, \tag{6}$$

$$E_{ij,k,\mathrm{B}} = \exp -\frac{(I_{ij,k,\mathrm{B}} - 0.5)^2}{2\sigma^2}, \tag{7}$$

$$E_{ij,k} = E_{ij,k,\mathrm{R}} \times E_{ij,k,\mathrm{G}} \times E_{ij,k,\mathrm{B}}, \tag{8}$$

ここで，σ は Mertens らの実装では 0.2 と設定されている．次に，これら 3 つの指標の単純な乗算によりスカラー値の重みマップ W を求める．線形和の重みのようにこれら 3 つの指標に乗数を用いた重み付けを行うと W は次式で求められる．

$$W_{ij,k} = (C_{ij,k})^{\omega_C} \times (S_{ij,k})^{\omega_S} \times (E_{ij,k})^{\omega_E}, \tag{9}$$

この式で，ω_C，ω_S，ω_E はそれぞれの項の相対的な重み付けのための係数である．標準値として $\omega_C = 1.0$，$\omega_S = 1.0$，$\omega_E = 1.0$ が用いられるため本研究では同様の値を用いる．最後に，合成時に重み付き平均を用いているため次式により $W_{ij,k}$ を正規化した $\hat{W}_{ij,k}$ を求める．

$$\hat{W}_{ij,k} = \frac{W_{ij,k}}{\sum_{k=1}^{N} W_{ij,k}}. \tag{10}$$

これらの処理により算出した合成重み $\hat{W}_{ij,k}$ を式 (2) の重みとし画像を合成する．

算出した重みを用いた画像合成にはラプラシアンピラミッドという複数解像度の画像ピラミッドを用いた合成方法を用いる．単純な重み付き合成では，重みにより異なる画像を用いている境目が目立ってしまい自然な画像が作成できない．そこで，ラプラシアンピラミッドを用いた画像合成を用いてより自然な合成画像を得る．

ラプラシアンピラミッドは，画像のガウシアンピラミッドを用いて作成される画像ピラミッドである [30]．まず，入力画像の高さと幅を 1/2 にするダウンサンプリングを l 回かけて作成したガウシアンピラミッドを作成する．この時作成したガウシアンピラミッドが持つ画像の枚数は $l + 1$ 枚である．次に，l 番の画像の画素値から $l - 1$ 番の画像をアップサンプリングした画像の画素値を引き差分を求める．この引算を入力画像である $l = 1$ 番目の画像まで行った差分の画像群がラプラシアンピラミッドである．図 5 にラプラシアンピラミッドの例を示す．

図 5　ガウシアンピラミッドとラプラシアンピラミッドの例．図中の L1 および L2 は画像特徴の視認性を高めるために元画像の各画素値を 16 倍にした画像．

画像の l 段階のラプラシアンピラミッドを作成する関数を $\mathrm{L}\{x\}^l$，l 段階のガウシアンピラミッドを作成する関数を $\mathrm{G}\{x\}^l$，x は任意の値を持つ画素値とすると合成処理は次のように表される．

$$\mathrm{L}\{R\}^l_{ij} = \sum_{k=1}^{n} \mathrm{G}\{\hat{W}\}^l_{ij,k} \mathrm{L}\{I\}^l_{ij,k}, \tag{11}$$

ここで，$\mathrm{L}\{R\}^l_{ij}$ は合成画像のラプラシアンピラミッドを表している．このラプラシアンピラミッドを作成した時と逆の処理を行うことにより合成画像が得られる．

2.5　機械学習

機械学習は，人工知能技術の一つであり，入力するデータからその特徴や傾向を学習し，固定的なプログラムでは解決できない分類や回帰などの問題に取り組む技術である [31]．次節で述べるニューラルネットワークおよび深層学習も機械学習手法の一種である．本節では，機械学習で最も一般的な教師あり学習とその手法について説明する．機械学習は，画像認識や物体認識などのコンピュータービジョン，ショッピングサイトのレコメンデーションシステムや，コンピュータウイルスの検出，ビッグデータの分析など現

在の生活に欠かせない技術である．

機械学習では，主に教師あり学習と教師なし学習の 2 種類の学習方法がある [31]．その他にも，近年では半教師あり学習や自己教師あり学習 [32]，敵対的学習などの学習方法も存在する [31]．2 つの学習方法の違いは大まかにいうと入力データと対になる教師データが存在するかしないかである．教師あり学習は画像認識に代表されるような分類や回帰といったタスクに用いられ，教師なし学習はデータから似た性質を持つものを集めるクラスタリングなどに用いられる．ここでは基本的な学習方法である教師あり学習について説明を行う．

2.5.1　教師あり学習手法

教師あり学習は，学習時に入力データと所望の出力を表す正解データを用いて学習する方法である．あらかじめ，入力データとその対になる正解データ（教師データ）を用いて学習するため「教師あり」学習と呼ばれる．教師あり学習を用いる手法にはサポートベクターマシーン（Support Vector Machine: SVM）や，k 近傍法，後述するニューラルネットワークおよび深層学習など様々な手法がある [33]．ニューラルネットワークおよび深層学習は 2.6 節で説明するため，ここでは SVM と k 近傍法について簡単に説明する．

SVM は，非線形な分類問題にも対応した教師あり機械学習アルゴリズムである [33]．初めて提案された SVM は線形分離可能なデータに適用できる線形 SVM であり，後に非線形のデータにも対応できる非線形 SVM が提案された．SVM は，非線形空間における分類問題において凸最適化により学習結果が一意に定まる点や簡単にプログラミングで利用できるライブラリが公開されている点などが主な特徴である．簡単な SVM として 2 クラス分類問題における線形 SVM について述べる．線形 SVM は，入力した特徴データからそれを各クラスに分離する関数である線形分類関数を用いる．SVM では，その識別関数から最も近い学習データからその関数までの距離を最大化することにより分類の学習を行う．非線形 SVM では非線形の分類を可能とするためにカーネルトリックと呼ばれる手法が用いられる．

k 近傍法は，特徴空間における各特徴量間の距離の近さにより分類を行う単純な手法である [33]．未知のデータに対して，そのデータに特徴空間上で近い学習データの K 個分のデータからその未知データのクラスの確率を算出し分類を行う．具体的には，下記の式により確率を計算する．

$$p(y = c|\boldsymbol{x}, \mathcal{D}, K) = \frac{1}{K} \sum_{i \in N_K(\boldsymbol{x}, \mathcal{D})} \mathbb{I}(y_i = c), \tag{12}$$

$$\mathbb{I}(e) = \begin{cases} 1 & \text{if } e \text{ is True} \\ 0 & \text{if } e \text{ is False} \end{cases}, \tag{13}$$

ここで，$N_K(\boldsymbol{x}, \mathcal{D})$ は K 個の $\boldsymbol{x}$ に近い $\mathcal{D}$ 空間でのデータのインデックス集合である．k 近傍法は，単純な分類アルゴリズムであるが，あるクラスのデータが 1 個からでも分類の学習が可能であるという特徴がある．

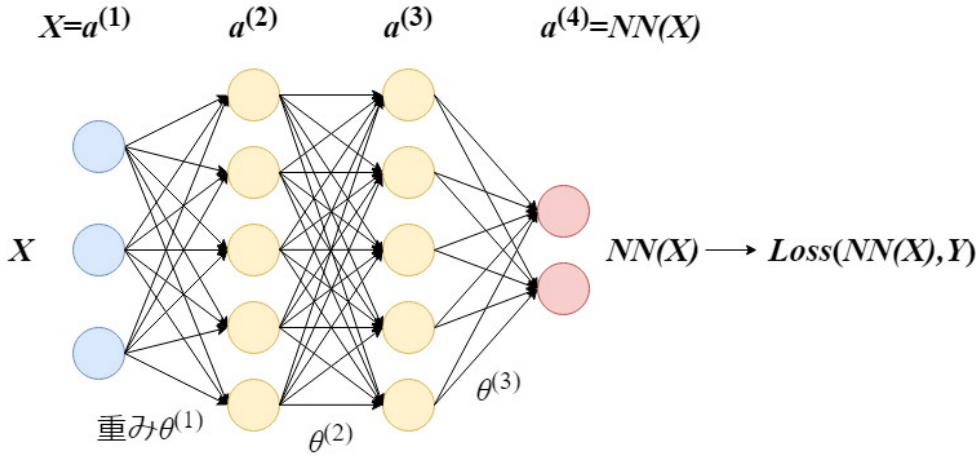
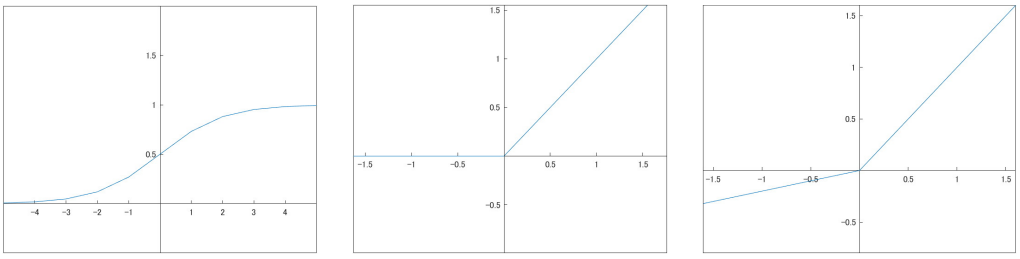


図 6　全結合層を用いた 4 層のニューラルネットワークの例

2.6　ニューラルネットワークおよび深層学習

ニューラルネットワークは，人の脳の神経細胞ネットワークの動作から着想しモデル化した機械学習手法である [31]．人の脳の神経細胞であるニューロンは，ほかのニューロンから信号を受け取り，その信号のパターンから次のニューロンにシナプスを通して信号を伝達する働きをする．ニューラルネットワークは，のニューロンの動作を入力信号の重み付き線形和と非線形関数によりモデル化し，それを多数つなげたネットワークとして構成したモデルである．その非線形関数は活性化関数と呼ばれる．前述したモデルは非常に簡単なモデル化であるため，より実際のニューロンやシナプスに伝わる信号の動作に近くなるようモデル化したニューラルネットワークも提案されている [34]．一般的には，入力から出力まで階層構造を持つネットワークに構成したものが用いられる．階層を持つニューラルネットワークでは，入力データを受け取る入力層と結果を出力する出力層，入力層と出力層の間にある中間層という層を持つ．深層学習は，この中間層を多段に接続して「深い」層構造を持たせたニューラルネットワークを用いることからそのように命名されている．

ニューラルネットワークの学習は，各層の重み付き線形和の重みを誤差逆伝播法と勾配降下法を用いて決定するのが一般的である [31, 33]．誤差逆伝播法は，ニューラルネットワークの各層構造を多項式の合成関数として表現できることを利用して，出力と正解データとの誤差を表す損失関数の値から各層の重みの更新を行う手法である． ここでは，図 6 に示す単純な 4 層のニューラルネットワークの場合を考え学習方法を説明する．図 6 において，入力データを $\boldsymbol{X}$，各層の入力と重みを $\boldsymbol{a}^{(i)}, \theta^{(i)}$，$i \in 1, 2, 3, 4$，教師となる正解データを $\boldsymbol{Y}$ とする．図 6 のニューラルネットワークは，活性化関数として Logistic Sigmoid 関数を用いるとする．本論文では，Logistic Sigmoid 関数を単に Sigmoid 関数と



（a） Sigmoid 関数　　（b） ReLU 関数　　（c） Leaky ReLU 関数

図 7　代表的なニューラルネットワークの活性化関数

呼ぶ．Sigmoid 関数は次式で表される．

$$\mathrm{Sig}(x) = \frac{1}{1+\exp(-x)}. \tag{14}$$

また，ここで損失関数は以下の二乗誤差を用いると考える．

$$Loss(\boldsymbol{x}, \boldsymbol{y}) = \frac{1}{2}(\boldsymbol{x} - \boldsymbol{y})^2. \tag{15}$$

まず，入力 $\boldsymbol{X}$ をネットワークに入力した時の各層の出力は以下のように計算できる．

$$\boldsymbol{a}^{(1)} = \boldsymbol{X}, \tag{16}$$
$$\boldsymbol{a}^{(2)} = \mathrm{Sig}(\theta^{(1)}\boldsymbol{a}^{(1)}), \tag{17}$$
$$\boldsymbol{a}^{(3)} = \mathrm{Sig}(\theta^{(2)}\boldsymbol{a}^{(2)}), \tag{18}$$
$$\boldsymbol{a}^{(4)} = \mathrm{Sig}(\theta^{(3)}\boldsymbol{a}^{(3)}), \tag{19}$$
$$\mathrm{NN}(\boldsymbol{X}) = \boldsymbol{a}^{(4)}, \tag{20}$$
$$Loss(\mathrm{NN}(\boldsymbol{X}), \boldsymbol{Y}) = \frac{1}{2}(\mathrm{NN}(\boldsymbol{X}) - \boldsymbol{Y})^2, \tag{21}$$

ここで，ニューラルネットワークの各重みを更新には勾配降下法を用いるため，損失関数を各重みで偏微分した値 $\frac{\partial Loss}{\partial \theta^{(i)}}$ が必要になる．それら偏微分値を求める時には，各層の出力が合成関数として表現できることを利用し合成関数の微分法を用いて各重みの偏微分値を求める．損失関数の各重みによる偏微分値は次式のように算出される．

$$\frac{\partial Loss}{\partial \boldsymbol{a}^{(4)}} = NN(X) - Y, \tag{22}$$
$$\begin{aligned}\frac{\partial Loss}{\partial \theta^{(3)}} &= \frac{\partial Loss}{\partial \boldsymbol{a}^{(4)}}\frac{\partial \boldsymbol{a}^{(4)}}{\partial \theta^{(3)}} \\ &= (\theta^{(3)})^{\mathsf{T}}\frac{\partial Loss}{\partial \boldsymbol{a}^{(4)}} \odot \mathrm{Sig}'(\theta^{(3)}\boldsymbol{a}^{(3)}),\end{aligned} \tag{23}$$
$$\begin{aligned}\frac{\partial Loss}{\partial \theta^{(2)}} &= \frac{\partial Loss}{\partial \boldsymbol{a}^{(3)}}\frac{\partial \boldsymbol{a}^{(3)}}{\partial \theta^{(2)}} \\ &= (\theta^{(2)})^{\mathsf{T}}\frac{\partial \boldsymbol{a}^{(4)}}{\partial \theta^{(3)}} \odot \mathrm{Sig}'(\theta^{(2)}\boldsymbol{a}^{(2)}).\end{aligned} \tag{24}$$

誤差逆伝播法は，正解データと出力から損失関数を用いて算出された誤差値を基に，出力の計算時とは逆順で各重みの偏微分値を求めていくためそのように呼ばれる．

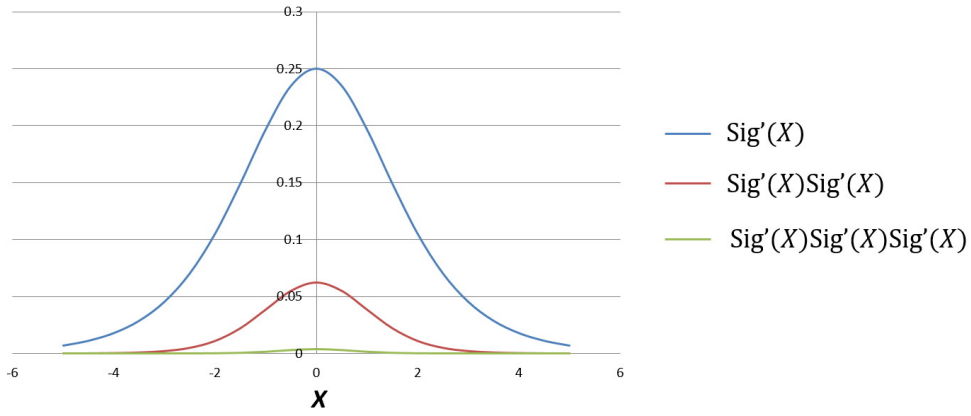


図 8　Sigmoid 関数の微分値の積

ニューラルネットワークには，多段構造の場合において学習時に勾配消失問題がしばしば発生する [35]．勾配消失問題の原因の 1 つに活性化関数があげられる．ここで前述した簡単なニューラルネットワークで用いている Sigmoid 関数の微分値について考える．Sigmoid 関数の微分値は次式で求められる．

$$\mathrm{Sig}'(X) = \frac{d}{dX}(\mathrm{Sig}(X)) = \mathrm{Sig}(X)(1 - \mathrm{Sig}(X)). \tag{25}$$

前述した合成関数により，各層で算出される偏微分値は Sigmoid 関数の微分値の積になる．図 8 に Sigmoid 関数の微分値のグラフとさらに乗算していった式のグラフを示す．グラフより，微分値を乗算するほどその値が小さくなることがわかる．つまり，入力層に近くなるにつれ勾配を求める時に乗算の回数が増えその値が小さくなり，深い層構造を持つニューラルネットワークでより顕著になる．その非線形関数の偏微分値が小さくなることが大きな原因になり，ニューラルネットワークでは学習時に伝播していった勾配が徐々に消失し学習に失敗する勾配消失問題が発生する．勾配消失問題は，ニューラルネットワークの層が多段になればなるほど顕著になる．それに対して Rectified Linear Unit（ReLU）関数を用いることでその問題を回避しする手法が提案され応用における性能向上に寄与している [36]．ReLU 関数は下記の式で表される非線形関数である．

$$y = max(0, x). \tag{26}$$

図 7 に，Sigmoid 関数と ReLU 関数，LeakyReLU 関数 [37] のグラフを示す．図 7 では，視覚的に違いを分かりやすくするため LeakyReLU 関数の入力値の 0 以下のグラフの傾きを決めるパラメータを 0.2 にしている．Hinton らによって提案された ReLU 関数は，Fukushima らの論文で既に示され活性化関数として用いられているものと同じものでもある [38]．

深層学習は，従来よりも層数が多く深いニューラルネットワークを用いた機械学習手法である [31]．深層学習は，それまで機械学習を用いた画像認識手法で分割した 2 つの処

理であった特徴量抽出とその分類を同一のニューラルネットワーク上で行うという特徴がある．特に，2012 年の画像認識コンペティションでは，フィッシャーベクター特徴量と SVM を用いた従来の機械学習手法に比べ 10 %以上の認識精度向上を達成した [39]．深層学習は，多量のデータから多段の層にある非常に大量なパラメータの学習を行うため，その学習に計算量が必要となる．学習では，テンソルの畳み込みや積といった演算が行われるため，並列計算能力に優れたグラフィックスプロセッシングユニット（GPU）を用いた計算を行うのが一般的である．その他にも，テンソルの計算に特化したプロセッサ [40] や深層学習のためのプログラミング用ライブラリ [41] なども数多く開発されている．

深層学習の学習では，損失関数が非凸関数となるためミニバッチ確率的勾配降下法を用いて学習を行う．ミニバッチ確率的勾配降下法とは，学習に用いるデータをミニバッチと呼ばれる細かいデータの集まりに分け，学習するパラメータの勾配計算と値の更新をミニバッチごとに行う手法である．深層学習やニューラルネットワークは，非線形関数と複雑なネットワーク構造により非線形性を有しており，その損失関数は非凸関数になる．SVM やその他のアルゴリズムで用いられるような収束が保証された凸最適化アルゴリズムは用いず，単純な勾配を用いたミニバッチ確率的勾配降下法を用いるのが一般的である．確率的勾配降下法を非凸関数の最適化に用いるため，深層学習の最適化においては収束の保証はない [31]．

ミニバッチ確率的勾配降下法には様々な手法が提案されている [42–44]．Momentum SGD 法は，確率的勾配降下法の学習を高速化する目的で提案された手法である [42]．Momentum SGD 法は，過去の勾配の移動平均を蓄積しそれを加味して継続的にその勾配の方向に進むような更新値が設定される．Momentum という名称からも分かるように，物体が運動したときに加わる慣性のような働きを追加しているアルゴリズムである．

AdaGrad 法は，設定する学習率のパラメータを過去の勾配の二乗和の平方根で割ることで，各勾配の値ごとに適応的に学習率を設定する手法である [43]．学習する変数を x，損失関数の勾配を g_t としたとき，t 回目の更新で AdaGrad 法により計算される更新量 Δx_t は以下の式で算出される．

$$\Delta x_t = -\frac{\eta}{\sqrt{\sum_{\tau=1}^{t} g_\tau^2}} g_t, \tag{27}$$

ここで，τ は学習率であり任意に設定されるハイパーパラメータである．

ADADELTA 法は，AdaGrad 法では手動で設定が必要だったグローバルな学習率のハイパーパラメータ設定を自動化し，また AdaGrad 法では蓄積した過去全ての勾配情報を用いるため学習率に過去の勾配の影響が残り続けるという問題がありその解決も行うために提案された手法である [44]．ADADELTA 法は，過去の勾配情報に対して指数関数的に減衰をかけたものを計算に使用する．学習する変数を x としたとき，ADADELTA

法の更新量の計算式は次式である．

$$E[g^2]_t = \rho E[g^2]_{t-1} + (1-\rho)g_t^2, \tag{28}$$

$$E[\Delta x^2]_t = \rho E[\Delta x^2]_{t-1} + (1-\rho)E[\Delta x^2]_t, \tag{29}$$

$$\Delta x_t = -\frac{\sqrt{E[\Delta x^2]_t + \varepsilon}}{\sqrt{E[g^2]_t + \varepsilon}} g_t, \tag{30}$$

ここで，ρ は一定の値を持つパラメータであり過去の値の減衰度合いを決定し，ε は分数の 0 割りを回避するための定数である．学習する変数とその更新のための値は同じ単位であるべきであるが，式 (27) の AdaGrad 法では更新に使う値が無単位量となる．そこで，ADADELTA 法では式 (30) により更新量の単位を揃えたことで，単位の違いを吸収していた学習率のハイパーパラメータを更新値の計算から無くすことができ，学習率の設定が必要なくなった．

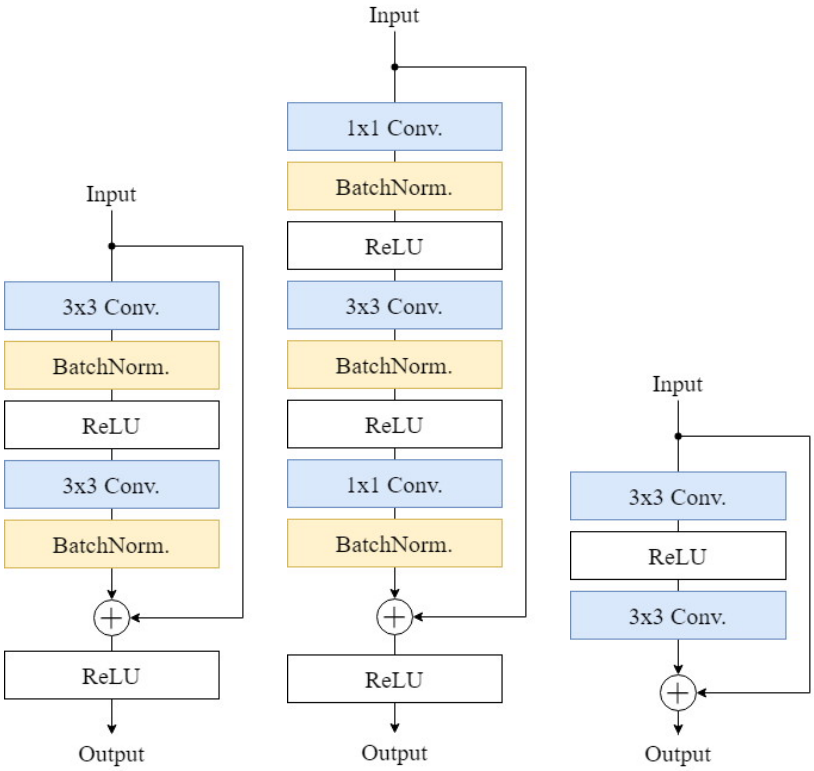
2.6.1　畳み込みニューラルネットワーク

CNN は，畳み込み層と呼ばれる畳み込み演算を行う層を組み込んだニューラルネットワークの一種であり，画像分類や物体検出などのコンピュータービジョンと画像補間や画像生成などの画像処理手法で非常によい成果をあげている [38, 39, 45–57]．一般的に CNN は，畳み込み演算を行う畳み込み層と活性化関数，入力信号の間引きを行うプーリング層などを複数層組み合わせて構成されている．畳み込み層は，入力信号に対して学習可能な係数を持つフィルタを一定間隔でスライドさせながら畳み込む演算を行い，高さ h および幅 w，チャンネル数 c の要素を持つ三次元信号を出力する．本論文では，そのスライドさせる幅をストライドと呼び，畳み込み層の出力を特徴マップと呼ぶ．2 次元の畳み込み層では，前述した大きさの三次元信号に対し $h_{\mathrm{filter}} \times h_{\mathrm{filter}} \times c_{\mathrm{in}}$ の大きさを持つフィルタを c_{out} 個用意してそれぞれ畳み込む．ここで，h_{filter} は畳み込むフィルタの高さと幅の大きさであり，c_{in} と c_{out} はフィルタのチャンネル方向の大きさと個数である．それぞれの畳み込み演算の結果を画像のチャンネル方向に結合したものが畳み込み層の出力特徴マップとなる．

CNN の活性化関数には，ReLU 関数が一般的に用いられる [36]．一般的に，CNN モデルの出力における一要素を算出するために用いられる入力画像の画素集合を受容野と呼ぶ．受容野の大きさは，CNN モデルの構造に依存し，その大きさが大きいほど入力画像の広い領域を畳み込んで出力結果を算出することになるため，受容野が大きいほど画像内の広い領域にわたる特徴が学習できると言われている [53]．CNN モデルのパラメータ学習は，通常のニューラルネットワークと同様に誤差逆伝播法を用いて学習する [45]．

2.6.2　Residual Block

本項では，画像認識や超解像などで高い成果を上げているニューラルネットワークの構造である Residual block（ResBlock）について述べる [49, 52]．図 9（a），（b）および



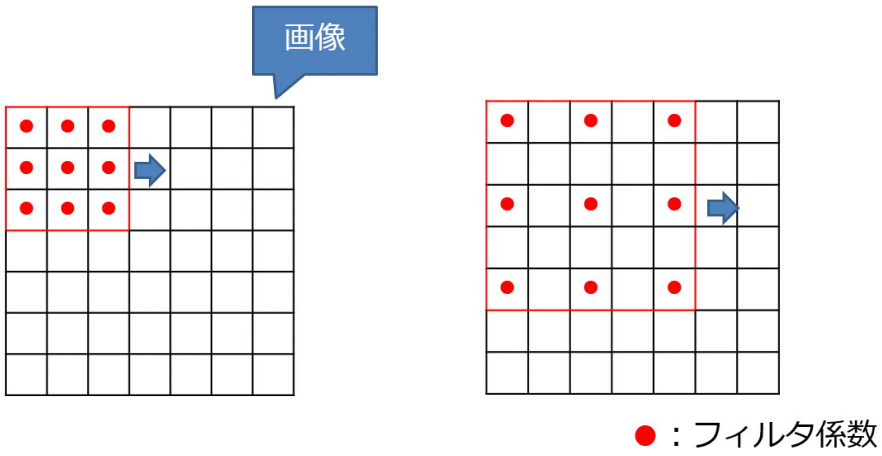
（a）ResBlock [49]　（b）Bottleneck [49]　（c） [52]

図 9　ResBlock の構造

（c）に画像認識手法で提案された ResBlock [49] と提案法で用いる画像超解像手法 [52] で提案された ResBlock の構造を示す．ここで，“Conv.” は畳み込み層を “Batch Norm.” は Batch Normalization [58] を示している．ResBlock は，図 9 に示すように 2 層または 3 層の畳み込み層と活性化関数，SkipConnection と呼ばれる入力特徴マップを出力との要素ごとの和を算出する構造から構成されている．SkipConnection により入力と出力の恒等変換を学習が可能になる．さらに，SkipConnection があるために誤差逆伝播法により勾配計算時に，より入力に近い層にその勾配情報を伝えることが容易になる．それらの効果により，モデル学習時に発生する勾配消失問題 [35] を軽減しモデルの層数を従来に比べ飛躍的に増加させることを可能とした構造である．ResBlock を多段接続したモデルは層数が 100 層以上あっても学習可能になることが知られている [49]．ResBlock は，そのモデルが学習するタスクにより様々な種類の構造が提案されている．本論文では，図 9（c）に示す画像超解像手法で提案された Batch Normalization を取り除いた ResBlock を基に提案 CNN モデルを構成した [52].

2.6.3　Dilated 畳み込み層

Dilated 畳み込み層は，通常の畳み込み層よりもモデルの受容野を広くでき，通常の畳み込み層に比べ入力信号の広い領域にわたる特徴を学習できる畳み込み層の 1 つであ



（a） 通常の畳み込み層　　（b）Dilated 畳み込み層

図 10　Dilated 畳み込み層と通常の畳み込み層のフィルタ畳み込み演算

る [59]．図 10 に，通常の畳み込み層でのフィルタ係数の適用方法と Dilated 畳み込み層での適用方法を示す．図 10 において，赤点がフィルタの各係数を表しており，マス目が入力画像の 1 画素を表している．図に示すように通常の畳み込み層は，フィルタ係数を入力の隣り合う要素に適用するのに対し，Dilated 畳み込み層では高さと幅の方向に一定間隔で離れた要素にフィルタを適用する．Dilated 畳み込み層の処理を式で表すと次のようになる．

$$\boldsymbol{y}_{i,j} = \sum_{u=-k'_h}^{k'_h} \sum_{v=-k'_w}^{k'_w} \boldsymbol{W}_{k'_h+u,k'_w+v} \boldsymbol{x}_{i+lu,j+lv}, \tag{31}$$

ここで，y は出力信号，x は入力信号，W はフィルタ係数，l フィルタ係数の間隔を決める定数である．本論文では，l を Dilation と呼ぶ．

$$k'_h = \frac{k_h - 1}{2}, \tag{32}$$

$$k'_w = \frac{k_w - 1}{2}, \tag{33}$$

この式で，k_h と k_w はフィルタカーネルの大きさを表す．この Dilation が，2 の場合は入力信号の一つ離れた要素毎にフィルタが適用され，3 の場合は二つ離れた要素毎に適用される．文献 [59] では，Dilation は構造内にある 1 層目の Dilated 畳み込み層から 2^n，（$n = 1, 2, 3, ...$）と変化するように設定されている．それにより，受容野を層数に対して 2 乗の指数関数的に拡大でき，ResBlock との組み合わせにより畳み込み層数が多く入力画像の広い領域にわたる特徴を学習できるモデルとなる．

2.6.4　Transposed 畳み込み層

Transposed 畳み込み層は，Deconvolution 層とも呼ばれ特徴マップの高さおよび幅の拡大に用いられる層の一つである [60]．Transposed 畳み込み層では，入力特徴マップの各要素の間に値が 0 の要素を挿入して拡張し，畳み込み層と同様の畳み込み処理を行う．これにより，出力される特徴マップは入力する特徴マップの高さおよび幅を拡大できる．拡大の大きさは、入力特徴マップの各要素間に挿入する 0 の要素の数とフィルタのスライド幅およびフィルタの大きさで決定される．

2.6.5　Contextual Attntion 層

Contextual Attntion 層は，画像補間手法の CNN モデルにおいて欠損領域の周辺画素から物体表面の模様のような詳細特徴を復元するための層である [57]．Contextual Attntion 層は，2 つの入力画像の特徴マップを用いる．ここでは，詳細情報の復元元の画像をソース画像，復元先の特徴マップをターゲット画像と呼ぶ．通常の画像補間であればソース画像が入力画像の欠損領域以外の画素になり，ターゲット画像が大まかに推定した欠損領域の画像情報となる．Contextual Attention 層の処理は，まず初めにソース画像を 3×3 のパッチ画像に分割する．次に，対象画像の 3×3 の領域 $t_{i',j'}$ とパッチ画像 $s_{i,j}$ とで正規化した内積を Softmax 関数で以下の式のように算出できる．

$$p_{i,j,i',j'} = \mathrm{softmax}\left(\langle \frac{\boldsymbol{s}_{i,j}}{\|\boldsymbol{s}_{i,j}\|}, \frac{\boldsymbol{t}_{i',j'}}{\|\boldsymbol{t}_{i',j'}\|} \rangle\right), \tag{34}$$

ここで，$p_{i,j,i',j'}$ は各画素の Attention スコアを示している．Attention スコアは 2 つの特徴マップ間で特徴が似ているかを示す．この処理は畳み込み層と特徴マップのチャンネル方向ごとの Softmax 関数で簡単に実装が可能である．最終的に，ソース画像のパッチ画像を Transposed 畳み込み層のフィルタの係数として用いて畳み込み演算を行う．それらの処理により 2 つの特徴マップ間で特徴が似たターゲット画像の領域にソース画像の詳細な特徴をコピーできる．

2.6.6　学習データの前処理および拡張

ニューラルネットワークや CNN などの機械学習手法の学習では，分類や回帰を行う機械学習手法で扱えるようにデータの前処理を行う．データの前処理としては，汎化性能を向上させるためのデータ拡張や学習モデルが考慮すべきデータの変動を抑える主に二つの処理が用いられる．学習用データの拡張では，例えば同じ物体を撮影した画像に画像処理によりノイズの付加や画像の切り取りなどを適用することで，様々なパターンの入力データを用意する．これにより，入力データ数とその種類を拡張し学習済みモデルの汎化性能を向上させられる．本節では，画像を入力とする CNN モデルの学習において用いられるデータの前処理および拡張方法について述べる．

（a）原画像　（b）切り取り　（c）左右反転

（d）90 度回転　（e）180 度回転　（f）270 度回転

図 11　画像の切り取りおよび水平方向反転と回転の例

一番よく用いられる画像データの拡張は，画像の回転と水平方向の反転，画像の切り取り処理である．図 11 にそれぞれの画像拡張の例を示す．CNN モデルの各畳み込み層では，3×3 のフィルタを画像の 1 画素ごとに適用していく．そのため，その畳み込み層で抽出する画像特徴はその画像内の位置によらずに検出ができる．しかし畳み込み層はその演算方法からその特徴の回転や大きさの変動には対応していない．よって，入力画像の回転と水平方向の反転のデータ拡張を加えることにより，入力画像の物体の回転や反転に対しても頑健なモデルを学習可能になる．

その次に使われる処理としては，画像の切り取り処理があげられる．切り取り処理では，学習用画像 1 枚から決まったサイズの画像を複数枚切り取ることで，データ数と種類の拡張を行える．画像のダウンサンプリングやアップサンプリングがしやすいなどの理由から，CNN モデルの入力としては画像の高さと幅の大きさが 256×256 や 512×512 などの 2 の累乗の数を持つパッチ画像として切り取る場合が多い．しかし，この拡張方法では入力画像全体にわたる特徴が切り取られてしまい，学習には不向きになってしまう場合も考えられる．よって，切り取り処理だけでなく画像の縮小なども合わせるなど学習手法で対象としている問題に合わせてデータ拡張を行う必要がある．

2.7　オプティカルフロー推定

オプティカルフロー推定は，連続した複数枚の画像間で画素ごとに同一物体の移動量を推定するコンピュータービジョンの手法である [27]．オプティカルフローは画素ごとに独立した物体の動きを推定できるため，物体追跡や画像の位置合わせ，動画の手ぶれ補正，動画のインターレース除去，動画のフレーム補間，動き補償による動画の符号化

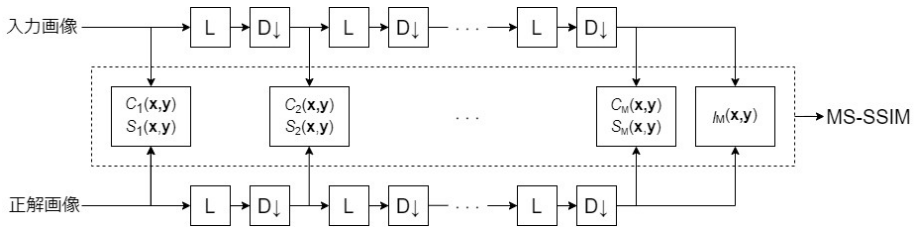


図 12　MS-SSIM の算出処理 [1]

などの技術に応用される．オプティカルフロー推定は主に 3 種類の手法があり，パッチ画像の類似度を用いる手法とオプティカルフローで画像変形した画像がもう一方の画像に近づくように正規化もしくは平滑化項を入れた最適化によりフローを推定する手法，CNN モデルを用いて画像の特徴抽出を行いその相関を用いてフローを推定する手法である [61–63].

CNN モデルを用いたオプティカルフロー推定手法で代表的なものは Flow-Net と PWC-Net である [61, 63]．Flow-Net は教師あり学習による学習を用いた CNN モデルによるオプティカルフロー推定手法である [61]．Flow-Net では，2 枚の画像を入力として直接オプティカルフローの推定結果を出力する 2 つのモデルを提案している．1 つ目のモデルは，畳み込み層と Transposed 畳み込み層により構成された単純な CNN モデルである．2 つ目のモデルは，Correlation 層と呼ぶ 2 つの画像から抽出した特徴マップをパッチごとに比較する層を組み込んだモデルである．それぞれのモデルを，GT のオプティカルフローのデータのあるデータセットにより End-to-End の教師あり学習を行い，学習させている．

2.8　画像評価指標

2.8.1　Peak Signal-to-Noise Ratio（PSNR）

画像処理の出力結果である画像の定量的評価には，一般的に PSNR[dB] が用いられる．PSNR は次式で計算される．

$$PSNR = 10\log_{10}\frac{255^2}{MSE}, \tag{35}$$

ここで，MSE は結果画像と正解画像における画素値ごとに計算した二乗誤差の平均値である．本論文では，真値となる正解画像が得られる場合に定量的評価指標として PSNR を用いる．

2.8.2　Multi-scale structual similarity（MS-SSIM）

Multi-scale structual similarity（MS-SSIM）は評価対象画像と真値画像との構造的類似度を算出する SSIM の拡張版であり，複数の画像解像度において構造的類似度を測定する

定量的評価指標である [1]．SSIM は，単一の解像度において画像の類似性を輝度項とコントラスト項および構造項の 3 つの項により類似度を算出する．それに対して MS-SSIM は，複数のダウンサンプリングを行いながら各解像度においてコントラスト項と構造項の値を算出し，それらの値と輝度項の値を用いて類似度を算出する．よって，SSIM よりも入力画像の解像度によらない画像の評価が可能である．本論文では，PSNR と同様に真値となる正解画像が得られる場合に定量的評価指標として MS-SSIM を用いる．

MS-SSIM の算出には図 12 のように複数回のダウンサンプリングを行いつつコントラスト項と構造項の値を算出していく．基本となる輝度項 l とコントラスト項 c および構造項 s は次式で算出される．

$$l(\boldsymbol{x},\boldsymbol{y}) = \frac{2\mu_x\mu_y + C_1}{\mu_x^2 + \mu_y^2 + C_1}, \tag{36}$$

$$c(\boldsymbol{x},\boldsymbol{y}) = \frac{2\sigma_x\sigma_y + C_2}{\sigma_x^2 + \sigma_y^2 + C_2}, \tag{37}$$

$$s(\boldsymbol{x},\boldsymbol{y}) = \frac{\sigma_{xy} + C_3}{\sigma_x\sigma_y + C_3}, \tag{38}$$

ここで，$\boldsymbol{x}$ および $\boldsymbol{y}$ は入力画像と正解画像の同一の位置から得られたパッチ画像である．また，$C_1 = (K_1L)^2$，$C_2 = (K_2L)^2$，$C_3 = C_2/2$ である．L は画像のダイナミックレンジを表す値であり 8bit 画像であれば $L = 255$ となる．また，K は $K_1 << 1$，$K_2 << 1$ のスカラー値である．これらの 3 つの項から SSIM の値は次式で計算される．

$$SSIM = \{l(\boldsymbol{X},\boldsymbol{y})\}^{\alpha} \cdot \{c(\boldsymbol{X},\boldsymbol{y})\}^{\beta} \cdot \{s(\boldsymbol{X},\boldsymbol{y})\}^{\gamma}, \tag{39}$$

ここで，α, β, γ は重み付けのためのパラメータであり，文献 [1] では $\alpha = \beta = \gamma = 1$ と設定されているため本論文でも同様の値を用いる．

複数解像度に対応させるため，図 12 に示すように M 段階で解像度を落としつつコントラスト項と構造項を算出する．図 12 において，L はローパスフィルタを表しており D は画像の大きさを $1/2$ にするダウンサンプリングである．図 12 より MS-SSIM は次式で表される．

$$MS\text{-}SSIM(\boldsymbol{x},\boldsymbol{y}) = \{l_{\mathrm{M}}(\boldsymbol{x},\boldsymbol{y})\}^{\alpha_M} \cdot \prod_{j=1}^{M} \{c_j(\boldsymbol{x},\boldsymbol{y})\}^{\beta_j} \{s_j(\boldsymbol{x},\boldsymbol{y})\}^{\gamma_j}, \tag{40}$$

ここで，$\alpha_j = \beta_j = \gamma_j$ であり，$\sum_j^M \beta_j = 1$ となるように各値が決められる．また，各 γ_j の値は複数解像度のなかで中間の解像度の画像が重視されるようにガウス分布を用いて算出される．

2.8.3　High Dynamic Range Visual Difference Predictor 2（HDR-VDP2）

HDR-VDP2 は，Mantiuk らによって提案された HDR 画像の定量的評価指標であり，人の視覚特性を考慮して HDR 画像の評価を行う HDR-VDP を発展させた評価指標であ

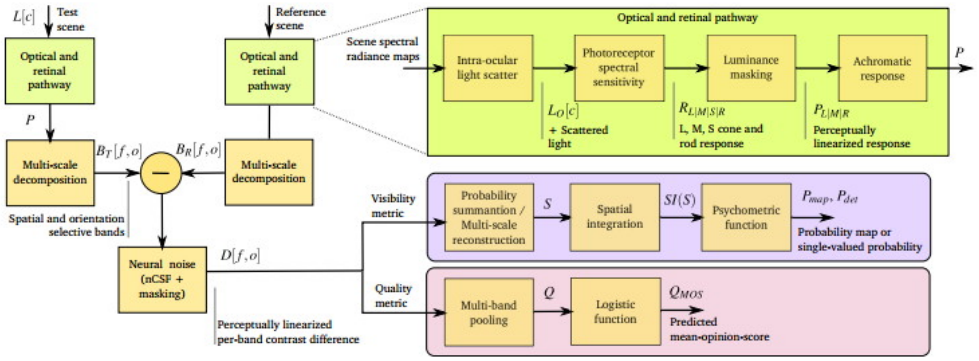


図 13　HDR-VDP2 の Q_{MOS} 算出処理フロー．文献 [2] の図 2 より引用

る [2]．HDR-VDP2 は，真値がある場合の HDR 画像の定量的評価指標として用いられている手法である．HDR-VDP2 は，HDR 画像の視認性（Visibility）と品質（Quality）で評価する．この 2 つのうち品質の指標 Q_{MOS} が，真値となる HDR 画像と評価対象の HDR 画像の 2 枚を用いて算出され定量的評価指標として用いられる．本論文では，HDR-VDP2 のスコアとして Q_{MOS} の値を用いる．Q_{MOS} は，0 から 100 の値であり値が高いほど良い画像であることを示す．

HDR-VDP2 のスコアは，図 13 で示す処理により算出される [2]．図 13 は，文献 [2] に示されている HDR-VDP2 の処理フローである．HDR-VDP2 はまず図 13 の「Optical and retinal pathway」により画像情報はフーリエ変換により周波数ごとに分解され，各処理により光学的および目の眼球と光受容体の特性を考慮した画像情報を算出する．次に，Multi-scale deconposition では画像を 4 方向の steerable pyramid を用いて算出した画像情報を分解する [64]．真値画像と測定対象画像の 2 つにおいて上記の処理を行い，その結果の差分値を算出する．最後に，差分値を用いて Q_{MOS} の値を算出している．

2.9　まとめ

第 2 章では，デジタルカメラの特性について説明し，多露光画像や HDR 画像，それらの合成手法や必要性を述べた．また，機械学習やニューラルネットワークや SVM など代表的な機械学習手法について述べ，CNN およびその関連技術，画像データの拡張方法について説明した．さらに，コンピュータービジョンの技術であるオプティカルフロー推定について述べ，画像処理において一般的に用いられる画像の画質評価指標について述べた．

第 3 章

関連研究

3.1 多露光画像のための画像内物体の位置補正手法

多露光画像のための画像内物体の位置補正手法は，多露光画像の内 1 枚を基準画像としてそれ以外の露光の画像の補正を行う手法である [3, 5, 14, 22, 23]. それらの手法では，画像内物体の位置やそのディテールも含め位置の補正を行う．従来は画像内のエッジやキーポイントと呼ばれる画像の特徴的な点を 2 つの画像間でマッチングして動きを推定し，補正を行う手法が提案されている．また画像を細かいパッチ画像に分け，画像特徴を用いて基準画像と同じ物体の位置の画像を再構成する PatchMatching 手法を用いて位置ずれを補正した多露光画像を作成する手法などが提案されている [23].

Tomaszewska と Mantiuk は，Scale Invariant Feature Transform（SIFT）特徴量を用いたマッチングによる多露光画像の補正手法を提案した [3]. その手法では，手持ち撮影による手ぶれなどにより発生する画像全体にわたる画像間の位置ずれを補正した多露光画像を生成することで合成時におけるアーティファクト抑制を行っている．図 14 は，文献 [3] に示されている Tomaszewska と Mantiuk が提案した補正手法の処理手順を示した図である．まず，適正露光に近い画像を基準画像としてそれ以外の画像を投影変換を用いた画像変形の対象とする．

次に，SIFT 特徴量 [65] を全ての露光の画像で計測する．SIFT 特徴量はその名前にもあるようにスケールと呼ばれる特徴量を算出する局所的な画像の範囲の大きさによらない画像の特徴量を抽出する手法である．SIFT 特徴量は，画像内から特徴点と呼ばれる画像内の座標において局所的な画像の輝度勾配の強度と方向から計算される特徴量である．この手法で測定される特徴点は，複数の大きさかつ複数の分散を持つガウシアンフィルタを用いて作成された複数解像度の画像群を用いて検出される．その後，特徴点が検出された解像度において特徴点の周辺 16×16 の画素における画像輝度の勾配方向および勾配の強度を計測する．次にその 16×16 領域を 4×4 の小領域に分割しその各領域ごとに 3 重の線形補間を用いて 8 個のヒストグラムを計算し，その $4 \times 4 \times 8 = 128$ 次元の非負ベクトルとして算出される．このベクトルの大きさを正規化することでさらに輝度変化に頑健な特徴量としている．

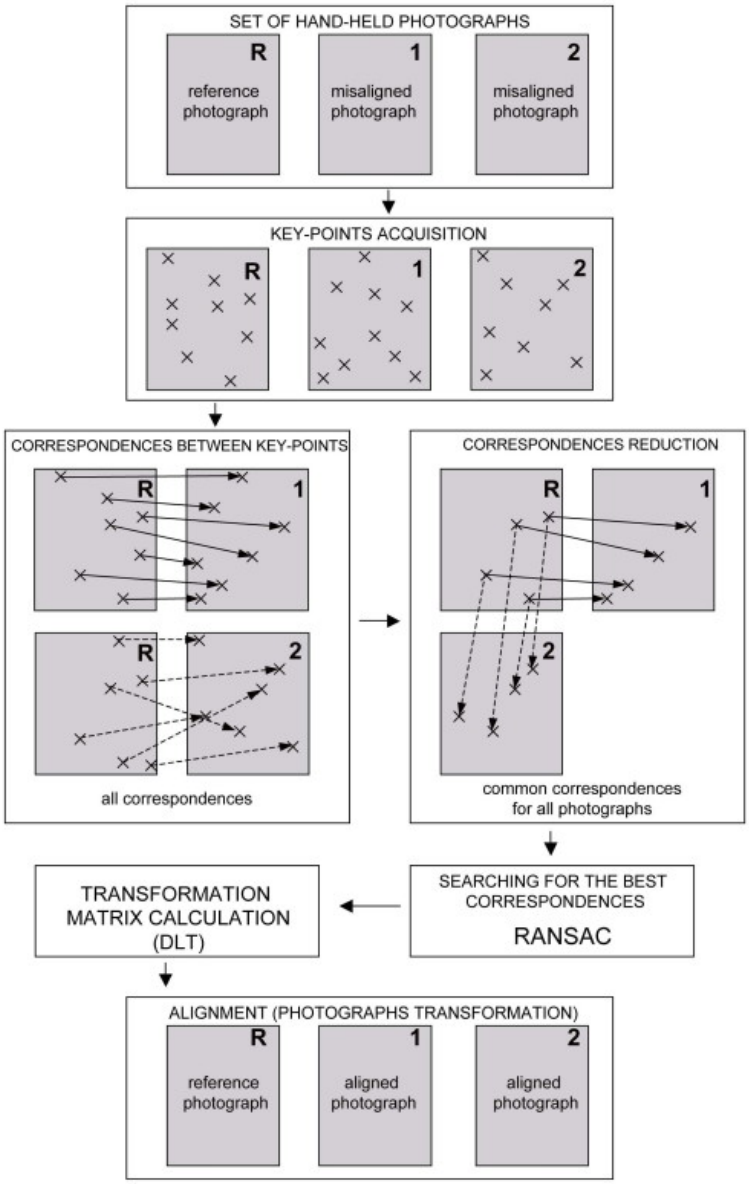


図 14　Tomaszewska と Mantiuk の画像補正手法の処理手順．文献 [3] の図 2 より引用

次に，基準画像とそれ以外の画像の間で各画像で計測された SIFT 特徴量を総当たりマッチングにより特徴点のペアを求め，RANSAC アルゴリズムと独自の基準によりより良い特徴点のペア選出する．画像全体のホモグラフィ変換には最低 4 個の特徴点のペアがあればよい．そのため，RANSAC アルゴリズムを用いて数ある特徴点のペアの中から同じアフィン変換のパラメータを示した特徴点のペアの数が多いもの 4 種類から，それらの代表となるペアのみを抽出しマッチングした特徴点のペアとする．この時，RANSAC アルゴリズムだけでは誤った特徴点のペアを算出する恐れがあるため，基準画像内の同じ特徴点が全ての他の露光画像の特徴点と同時にペアができているものを選ぶ．また，

抽出した特徴点のペアから Direct linear transform（DLT）を用いて変換行列 H を求め，

ホモグラフィ変換のための画像変形を行う．H は 3×3 の大きさの行列である．画像変形を行う時には画像に輝度値がない画素がある場合，バイリニア補間を用いて隣接画素値からその画素値を補間する．これにより画像全体の位置ずれを補正した多露光画像を作成している．

3.2　アーティファクト抑制を含む HDR 画像合成および多露光画像合成手法

多露光画像合成および HDR 画像合成では，アーティファクト抑制処理を含む多露光画像合成手法および HDR 画像合成手法が提案されている [6, 10, 11, 13, 16–21]．多露光画像合成では最新手法として Ma らの合成手法がある．また，HDR 画像合成では，Prabhaker らの CNN モデルを用いた Refinement を多露光画像に適用する手法と Niu らの Generative Adversarial Network（GAN）を用いた手法がある．

Ma らの手法は，多露光画像から適切な露光のパッチ画像を選択し，基準画像においてゴーストが発生しないと推定した領域に合成することでアーティファクトを抑制した合成画像を作成している [24]．まず，K 枚の多露光画像 $\{I_k, 1 \leq k \leq K\}$ から基準画像 I_r を選択し，その基準画像 I_r から画像の RGB ごとのヒストグラムマッチングを用いて $K-1$ 枚の画像群 $\{I'_k, k \neq r\}$ を作成する．基準画像に白とび黒つぶれが多い多露光画像の場合，$\{I'_k, k \neq r\}$ に画像内にはない色を持つ不自然なアーティファクトを発生させてしまう場合がある．次に，多露光画像 $\{I_k, 1 \leq k \leq K\}$ と基準画像から作成された画像群 $\{I'_k, k \neq r\}$ において各画像内から $N \times N$ の大きさを持つパッチ画像を抽出する．以降の式において，各画像のパッチ画像は全て同じ位置から抽出されたものとする．ここで，$\{I_k, 1 \leq k \leq K\}$ と $\{I'_k, k \neq r\}$ からパッチ画像を得るため次に，k 枚目の露光の画像から抽出されたパッチ画像の各画素値を一列に並べたベクトルを $\boldsymbol{x}_k$ および $\boldsymbol{x}'_k$ と表す．それぞれ，$\{I_k, 1 \leq k \leq K\}$ と $\{I'_k, k \neq r\}$ から抽出されたパッチから得られるベクトルである．$\boldsymbol{x}_k$ および $\boldsymbol{x}'_k$ はカラー画像から抽出されるため，入力が RGB 画像であれば $3N^2$ の大きさを持つ．そのパッチを次式を用いてそのパッチの信号強度と信号構造および平均輝度値の 3 つの要素に分解する．

$$
\begin{aligned}
\boldsymbol{x}_k &= \|\boldsymbol{x}_k - \mu_{\boldsymbol{x}_k}\|_2 \times \frac{\boldsymbol{x}_k - \mu_{\boldsymbol{x}_k}}{\|\boldsymbol{x}_k - \mu_{\boldsymbol{x}_k}\|_2} + \mu_{\boldsymbol{x}_k} \\
&= \|\tilde{\boldsymbol{x}}_k\|_2 \times \frac{\tilde{\boldsymbol{x}}_k}{\|\tilde{\boldsymbol{x}}_k\|_2} + \mu_{\boldsymbol{x}_k} \\
&= c_k \times \boldsymbol{s}_k + l_k,
\end{aligned} \tag{41}
$$

ここで，$\mu_{\boldsymbol{x}_k}$ は $\boldsymbol{x}_k$ 各要素の平均値であり，$\tilde{\boldsymbol{x}}_k$ は平均値を引いたパッチ画像である．また，c_k，$\boldsymbol{s}_k$，および l_k はそれぞれパッチ画像の信号強度と構造および平均輝度値を表す．

ここで，式 (41) の $\boldsymbol{s}_k$ の信号構造を I_r と $\{I_k, k \neq r\}$ から抽出されたパッチ画像で計算し，アーティファクトが発生する位置ずれがあるパッチか検出を行う．この検出で位

置ずれが検出された場合，$\boldsymbol{x}_k$ を $\boldsymbol{x}'_k$ と置き換えて以降の合成処理を行う．まず，基準画像から抽出されたパッチの $\boldsymbol{x}_r$ とそれ以外の画像の $\boldsymbol{x}_k$ から式 (41) を用いて $\boldsymbol{s}_r$ と $\boldsymbol{s}_k$ を求める．このベクトルが似ていればパッチ画像の構造的特徴が近く位置ずれがないと言えるため，$\boldsymbol{s}_r$ と $\boldsymbol{s}_k$ から構造的特徴の近さを表した ρ_k を次式により求める．

$$\rho_k = \frac{(\boldsymbol{x}_r - l_r)^T(\boldsymbol{x}_k - l_k) + \epsilon}{\|\boldsymbol{x}_r - l_r\|\|\boldsymbol{x}_k - l_k\| + \epsilon}, \tag{42}$$

この式で l_r は $\boldsymbol{x}_r$ の各要素の平均値であり，ϵ はセンサのノイズなどに頑健にするために可算される一定値であり，従来法では文献 [66] より $\epsilon = 0.00045$ としている．ρ_k を閾値処理し $\boldsymbol{x}_k$ と $\boldsymbol{x}'_k$ のどちらが用いられるか決定する．閾値処理は次式により表される．

$$\tilde{B}_k = \begin{cases} 1 & \text{if } \rho_k \geq 0.8 \\ 0 & \text{if } \rho_k < 0.8 \end{cases}, \tag{43}$$

$$\bar{B}_k = \begin{cases} 1 & \text{if } |l_k - l'_k| < 0.1 \\ 0 & \text{if } |l_k - l'_k| \leq 0.1 \end{cases}, \tag{44}$$

$$B_k = \tilde{B}_k \times \bar{B}_k, \tag{45}$$

ここで，$B_k = 1$ であれば $\boldsymbol{x}_k$ を用い，$B_k = 0$ ならば $\boldsymbol{x}'_k$ を用いるとしている．また閾値の 0.8 と 0.1 は従来法で定められた値である [24].

出力画像で使われるパッチ画像 $\hat{\boldsymbol{x}}$ は各露光のパッチ画像の値を用いて計算された $\hat{c}$, $\hat{\boldsymbol{s}}$，および $\hat{l}$ を用いて式 (41) の分解の逆処理をすることにより算出される．まず，c_k の信号強度は画像の局所的領域におけるコントラストと関係しており，高いコントラストを持つ画像は一般的に良い視認性を持っているとされる [24]．よって，次式を用いて複数露光のパッチ画像間で一番高い c_k を推定したパッチ画像の信号強度 $\hat{c}$ を算出する．

$$\hat{c} = \max_{1 \leq k \leq K} c_k. \tag{46}$$

信号構造 $\hat{\boldsymbol{s}}$ はベクトルであり，CN^2 次元空間での方向の情報を持っている．合成に用いられるパッチは各露光の信号構造を全て表現するものである方がよいため，$\hat{\boldsymbol{s}}$ は次式で算出される．

$$\hat{\boldsymbol{s}} = \frac{\bar{\boldsymbol{s}}}{\|\bar{\boldsymbol{s}}\|}, \tag{47}$$

$$\bar{\boldsymbol{s}} = \frac{\sum_{k=1}^{K} \|\tilde{\boldsymbol{x}}_k\|^p \boldsymbol{s}_k}{\|\tilde{\boldsymbol{x}}_k\|^p}, \tag{48}$$

$$p = 4.0, \tag{49}$$

ここで，$\|\tilde{\boldsymbol{x}}_k\|^p$ は各露光のパッチ画像の影響度合いを決める重み関数であり，画像の露光が高いものが選ばれるよう信号強度 c_k の値を用いて計算される．最後に，$\hat{l}$ はパッチ画像の平均輝度値の要素であり，より正確な色情報を持つ適正露光のパッチ画像の値を

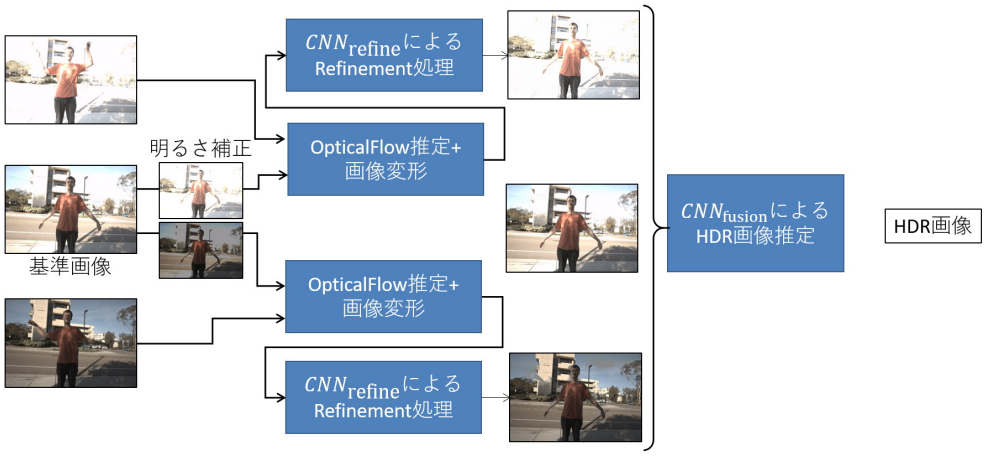


図 15　Prabhaker らの HDR 推定手法の概要

多く用いるように次式で算出される．

$$\hat{l} = \frac{\sum_{k=1}^{K} L(\mu_k, l_k) l_k}{L(\mu_k, l_k)}, \tag{50}$$

$$L(\mu_k, l_k) = \exp(-\frac{(\mu_k - 0.5)^2}{2\sigma_g^2} - \frac{(l_k - 0.5)^2}{2\sigma_l^2}), \tag{51}$$

ここで，μ_k は入力画像 I_k の平均輝度値であり，L は適正露光が多く撮られている多露光画像の内，画像全体と局所的な平均輝度値の 2 つの値からガウス分布を用いて決められる重み付け関数である．L により，多露光画像の中で画像全体の平均輝度値が 0.5 に近いかつ，局所的なパッチ画像内でも平均輝度値が 0.5 に近いパッチ画像の値が算出結果の多くを占めるように重み付けされる．最終的に，式 (41) から次式で合成パッチ画像を生成する [24].

$$\hat{\boldsymbol{x}} = \hat{c} \times \hat{\boldsymbol{s}} + \hat{l}. \tag{52}$$

最後に，画像内全てで計算された $\hat{\boldsymbol{x}}$ を用いて合成画像を出力している．

Prabhaker らの手法は，深層学習を用いたオプティカルフロー推定結果を用いた画像変形と CNN モデルによって推定した重みによる重み付き合成によって多露光画像の Refinement をし，さらに CNN モデルを用いて HDR 画像の合成を行う手法である [19]. 入力とする合計 3 枚の多露光画像 (I_0, I_1, I_2) の場合，その中間露光にあたる画像 I_0 を物体位置や姿勢の基準画像とする．選択した基準画像以外の画像 (I_1, I_2) が Refinement を行う対象の画像となる．ここで，(I_0, I_1, I_2) の画素値は，$[0, 1]$ の範囲に正規化されているとする．次に，選択した基準画像の明るさをその他の画像と同じになるよう補正する．明るさ補正は，次式を用いて処理を行う．

$$I_{0,j}(k) = \text{clip}(I_0(k) \times {\Delta_{0,j}}^{1/2.2}), \tag{53}$$

ここで，$\Delta_{0,j} = t_j / t_0$，$j = \{1, 2\}$ であり，$I_{0,j}(k)$ は I_j の画像を基準として明るさ補正を行った画像の k 番目の画素値，t_0 は基準画像の露光時間を t_j は j 番目の画像の露

光時間を表している．また，clip(x) は，x が $[0,1]$ の間の場合のみ値を保持し，0 以下の値は 0 に，1 以上の値は 1 にする処理である．明るさ補正を行った基準画像 $I_{0,j}$ と Refinement 対象画像 I_j，$j=\{1,2\}$ の間で，深層学習を用いたオプティカルフロー手法の PWC-Net [63] によりオプティカルフロー F_j を推定する．オプティカルフローは，2 枚の画像間で画素ごとに計測される移動量である．オプティカルフローの手法は画像の輝度変化に頑健な検出が行えるようそのモデル構造や学習に工夫が施されたモデルだが，より良い結果のために入力画像の明るさを揃えるのが一般的である．PWC-Net のニューラルネットワークモデルを PWC とすると F_j は次式で表される．

$$F_j = PWC(I_{0,j}, I_j). \tag{54}$$

次に，推定したオプティカルフローを用いて I_j，$j=\{1,2\}$ に対して画像変形を行う．画像変形により，画素値が無くなってしまう領域はバイリニア補間法を用いて周囲の画素情報から画素値を補間する．画像変形を Warp とすると変形後の画像は次式で表される．

$$I'_j = \mathrm{Warp}(I_{0,j}, I_j). \tag{55}$$

次に，オプティカルフローだけではそのエラーにより I'_j にアーティファクトが発生しているため，CNN モデルで生成した重みを用いた重み付き合成を用いて Refinement を行う．ここで，重み付き合成に用いられる画像は I'_j と $I_{0,j}$ である．Refinement 用の CNN モデルを CNN_{refine} とすると重み w は次式で表される．

$$W = CNN_{\mathrm{refine}}(I'_j, I_{0,j}, F_j). \tag{56}$$

そして，重み付き合成により Refinement された画像 $\hat{I}_j$ は次式で求められる．

$$\hat{I}_j = (1-W) \times I'_j + W \times I_{0,j}. \tag{57}$$

この Refinement 画像を基準画像を除く全ての露光の画像で作成する．

次に，推定用 CNN モデル CNN_{fusion} を用いて HDR 画像を推定する．CNN_{fusion} の入力画像には，$\{I_0, \hat{I}_1, \hat{I}_2\}$ が用いられ，推定 HDR 画像 $\hat{H}$ は次式で求められる．

$$\hat{H} = CNN_{\mathrm{fusion}}(I_0, \hat{I}_1, \hat{I}_2). \tag{58}$$

これらの CNN モデルを用いた処理により $\hat{H}$ を算出する [19]．これら 2 つの CNN モデルはどちらも U-Net [67] を基にした構造である．

Niu らの手法は，CNN モデルを用いて 3 枚の多露光画像から直接 HDR 画像を推定する手法であり CNN モデルの学習に GAN という深層生成モデルの学習方法を取り入れている [21]．まず入力する多露光画像 (I_0, I_1, I_2) をそれぞれガンマ補正した画像 (I_0^G, I_1^G, I_2^G) を用意し，それぞれの画像で色方向に結合したもの (X_0, X_1, X_2) を CNN モデルの入力として用いる．ここでも基準画像を I_0 とする．HDR 画像を CNN モデルにより直接推定するため，CNN モデルを CNN_{G} とすると推定 HDR 画像 $\hat{H}$ は次式で表される．

$$\hat{H} = CNN_{\mathrm{G}}(X_0, X_1, X_2). \tag{59}$$

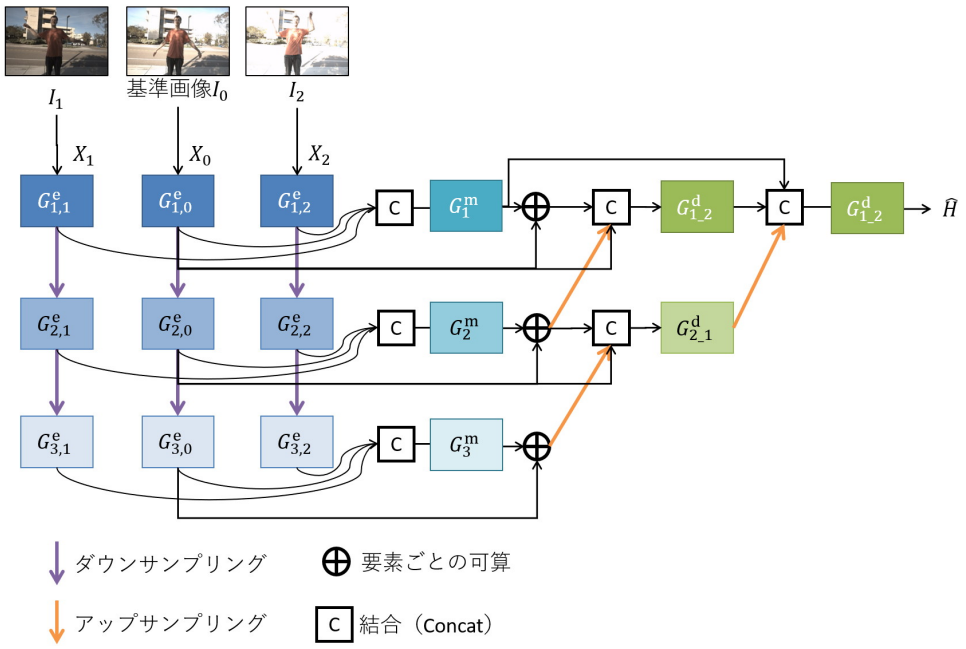


図 16　Niu らの CNN モデル CNN_{G} の概要

CNN_{G} の学習では，通常の正解画像と推定画像との平均絶対誤差に加えて推定を行う CNN_{G} 以外に Discriminator と呼ばれる CNN モデル CNN_{D} を用いてその 2 つの CNN モデル同士で敵対的にパラメータを学習する Adversarial 損失関数を用いた損失関数により学習している．

Niu らの手法は，3 種類の役割を持つ CNN モデルを組み合わせ構成されている．図 16 に CNN_{G} のネットワーク構造の概要を示す．入力 (X_0, X_1, X_2) はそれぞれ別の特徴抽出エンコーダネットワーク $G^{\mathrm{e}}_{j,i}$，（$j = \{1,2,3\}$，$i = \{0,1,2\}$）により 3 つの解像度で特徴を抽出した特徴マップを生成する．生成される特徴マップを式で表すと次式のようになる．

$$E_{1,i} = G^{\mathrm{e}}_{1,i}(X_i), \tag{60}$$
$$E_{2,i} = G^{\mathrm{e}}_{2,i}(F^{\mathrm{down}}(G^{\mathrm{e}}_{1,i}(X_i))), \tag{61}$$
$$E_{3,i} = G^{\mathrm{e}}_{3,i}(F^{\mathrm{down}}(G^{\mathrm{e}}_{2,i}(F^{\mathrm{down}}(G^{\mathrm{e}}_{1,i}(X_i))))), \tag{62}$$

ここで，F^{down} は特徴マップの高さと幅方向のダウンサンプリングを行う畳み込み層である．畳み込み層のフィルタは 3×3 の大きさでありストライドは 2 としている．次に，特徴マップ $E_{j,i}$ をそれぞれ特徴マップのチャンネル方向（画像を 3 次元テンソルとするとカラー方向）に結合し，結合 CNN モデル G^{m}_{j}，$j = \{1,2,3\}$ に入力する．G^{m}_{j}，$j = \{1,2,3\}$ により生成される特徴マップ M_j は次式で表される．

$$M_j = G^{\mathrm{m}}_{j}(\mathrm{Concat}(E_{j,1}, E_{j,0}, E_{j,2})) + E_{j,0}, \tag{63}$$

ここで，Concat は特徴マップの結合処理を示している．次に，特徴マップから HDR 画像を推定するデコーダ CNN モデル $G^d_{1\_1}$，$G^d_{2\_1}$，$G^d_{1\_2}$ を用いて推定 HDR 画像 $\hat{H}$ は次式で表される．

$$C_{1\_1} = G^{\mathrm{d}}_{1\_1}(\mathrm{Concat}(M_1, E_{1,0}, F^{\mathrm{up}}(M_2))), \tag{64}$$
$$C_{2\_1} = G^{\mathrm{d}}_{2\_1}(\mathrm{Concat}(M_2, E_{2,0}, F^{\mathrm{up}}(M_3))), \tag{65}$$
$$\hat{H} = G^{\mathrm{d}}_{1\_2}(\mathrm{Concat}(M_1, E_{1,0}, C_{1\_1}, F^{\mathrm{up}}(C_{2\_1}))), \tag{66}$$

ここで，F^{up} は特徴マップのアップサンプリングを行う層である．

3.3　画像補間 (Image Inpainting)

Image Inpainting は，画像内の欠損した領域を周囲の画素値から推定し，自然に修復する画像処理技術である [53,57,68–70]．従来の Image Inpainting 手法は，画像内の小領域を切り取ったパッチ画像と欠損した領域に隣接した画素の特徴とを比較し，隣接した領域に似た特徴を持つ画像領域を欠損領域に貼り付ける．それを欠損領域の周囲から徐々に欠損領域全体を埋めるように繰り返し行うことで，欠損領域を復元する手法が提案されている [69]．Criminisi らは，パッチ画像のテクスチャを用いたパッチ画像マッチング手法を用いた手法を提案した [69].

Image Inpainting の最新手法は，欠損領域の推定に教師あり学習の CNN や Generative Adversarial Network（GAN）を用いた深層学習手法である [53,57,70]．Pathak らは，初めて GAN を用いて大きな欠損領域の補間する手法を提案した [70]．Iizuka らは，二種類の Discriminator ネットワークを用いて画像全体および局所的な特徴の一貫性を考慮した手法を提案した [53]. また，Yu らは Contextual attention 層と呼ばれる新しい層と GAN を用いて細かな画像特徴も復元する手法を提案した [57].

3.4　本研究の位置づけ

本研究では，多露光画像合成および HDR 画像合成において発生するアーティファクトの抑制を目的とし，多露光画像の補正手法において高いアーティファクト抑制効果を持つ手法を実現するため新たな補正手法を提案する．従来研究においてアーティファクトの抑制には多露光画像の位置ずれを補正する処理手法と合成時にその抑制を考慮した方法で画像を作成する 2 つの方法が研究されてきた．最新研究では，合成時にアーティファクト抑制を考慮した方法が多く提案されておりパッチ画像や深層学習を用いたものが提案されている [17–21,24]．それらの手法では，多露光画像の基準画像とその他の露光画像間で発生する欠損領域は考慮されておらず，その欠損領域が発生した領域でアーティファクトを発生させている．よって，本研究ではその欠損領域を考慮した補間と多露光画像の画像間の位置合わせを行う多露光画像の補正手法を提案する．提案する補正

手法により位置合わせとアーティファクトの原因となっている欠損領域を補間した多露光画像を合成に用いることで，結果として合成画像のアーティファクトを抑制する．

本研究では，1 枚の既知の欠損領域を補間を行う一般的な画像補間とは異なり，複数枚の入力画像を持つ多露光画像のための欠損領域の検出と補間をする補正手法を提案する．従来の画像補間手法では，画像 1 枚の欠損領域は既知であるとしその 1 枚の入力から補間を行う処理や技術が提案されている．しかし，本研究で対象としている多露光画像の画像間での欠損領域は，その場所や大きさなどは既知ではなく画像ごとに異なりかつ多露光画像が持つ画像間での大きな露光の違いや白とび黒つぶれといった画素値の飽和によりその検出を容易にはできない．そこで，本研究では画像間の欠損領域を検出し 2 枚の入力画像からその欠損領域を補間する多露光画像のための補間手法を提案する．

最新研究では，HDR 画像合成や多露光画像合成に深層学習を用いたものが多く提案されている [17–21]．それら深層学習を用いた手法は教師あり学習によりニューラルネットワークのパラメータを学習しておりその多くは正解データとして従来の HDR 推定手法を用いて作成したものを用いている．そのため推定した HDR 画像は正確な情報の推定という点では，従来手法の精度以上にはほぼならなく，それらの正解画像は真値のデータであるとは厳密にはいえない．また HDR 画像と多露光画像合成手法では出力する信号が異なるため合成部分のモデルを共通に用いることは難しいと考えられる．対して多露光画像の位置ずれを補正する提案法では，合成の前処理として組み込むことができるため，結果的に合成処理からアーティファクト抑制処理部分を取り除くことができる．それにより HDR 画像の推定や合成画像の作成とアーティファクト抑制をそれぞれ別の手法で分けられるため，2 つの目標を同時に達成する研究を行う必要はなくそれぞれの目標の達成を目指して効率的に研究を進められると考えられる．さらに，厳密に真値といえる正解データ用の HDR 画像を用意するには一般的ではない HDR カメラで取得する必要があるなど非常に労力がいるが，位置ずれのない多露光画像は通常のカメラであれば取得可能であり，入力と正解データのデータセットの数を前者よりも容易に増やせる．よって本研究では利点が多く正解データの数が必要な深層学習を用いた教師あり学習手法を適用できる多露光画像の位置ずれ補正手法について研究を行った．4 章において，3.2 節で述べた従来の多露光画像補正手法や合成手法で対応できていない画素値飽和領域かつ動きによる欠損領域に対応したアーティファクト抑制する多露光画像補正手法を提案する．本研究の成果は，まず画素値飽和領域かつ動きによる欠損領域を定義したことである．さらに，5 章の実験においてその欠損領域の画像情報推定が，多露光画像合成および HDR 画像合成のアーティファクト抑制の前処理として効果があることを示した．これらの成果により従来よりもより効率的にアーティファクト抑制を行うことが可能になる．

3.5　3 章のまとめ

第 3 章では，従来のゴースト抑制を目的とした多露光画像合成手法および HDR 画像合成手法，最新の画像補間手法について紹介した．また，従来の多露光画像合成による合成

手法の問題点について述べ，本研究の位置づけを述べた．

第 4 章

深層学習を用いた多露光画像の位置ずれ補正手法

4.1 概要

本章では，本研究で提案する深層学習を用いた新たな多露光画像の位置ずれ補正手法について述べる．近年の従来のアーティファクト抑制手法は，多露光画像を合成する処理に抑制処理を加える手法が研究されてきた．しかし，アーティファクトの原因は多露光画像の画像間で発生している位置ずれによるものであり，位置ずれのない多露光画像であれば合成画像のアーティファクトは発生しない．よって本研究では，合成処理に抑制処理を加えるのではなく，位置ずれのある多露光画像から位置ずれのない多露光画像を作成する補正手法について研究を行う．多露光画像補正手法によるアーティファクト抑制は，抑制処理を含む多露光画像の合成手法に比べ次のような利点を持つ．

- 多露光画像の合成処理からアーティファクト抑制の問題を分離できるため，合成処理はより正確な合成画像の生成に注力できる．
- 多露光画像合成技術と HDR 画像合成技術のどちらに対しても応用できる．

この多露光画像補正の問題に対して，本研究では従来の補正手法に加え 2 つの提案 CNN モデルを用いた補正手法を提案する．従来のアーティファクト抑制手法では，合成画像のアーティファクトが特定の領域で発生していることが分かった．その領域とは，多露光画像の画像間の物体位置ずれと位置の基準画像内に発生した白とびおよび黒つぶれが同一の領域に発生し画像の情報が欠損した領域である．従来の補正手法および抑制処理を含む合成手法では，その欠損により位置ずれの推定や補正が十分にできず合成結果にアーティファクトを発生させていた．そこで，本研究では従来の位置ずれ補正に加え，その欠損領域を検出し欠損した画像情報を補間する手法を提案する．

図 17 に提案法と多露光画像合成手法を組み合わせた多露光画像の合成処理フローを示す．提案法は，合成画像のアーティファクト抑制を目的として，多露光画像のうち 1 枚を基準画像 I_{ref} とし，画像内物体やその詳細な特徴の位置合わせを行った多露光画像を

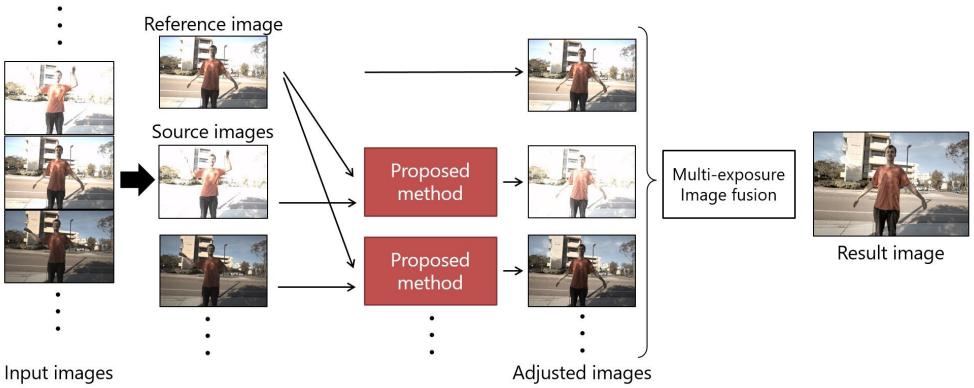


図 17　提案法を用いた多露光画像の補正と合成

生成する．まず初めに，基準画像 I_{ref} と補正を行う対象画像 I_{src} と呼ぶ．I_{src} は，例えば合計 3 枚の多露光画像 I_1, I_2, I_3 があり I_1 を基準画像 I_{ref} とした場合，I_{src} は I_2と I_3 となる．それぞれの I_2と I_3 に提案法を適用し，I_1 を基準として補正を行った画像を出力する．図 17 では，入力する多露光画像のなかで中間輝度の画像を I_{ref} としている．提案法によって出力された位置のあった多露光画像により，アーティファクトを抑制した合成画像が得られる．

提案法の補正処理は，従来法のオプティカルフローによる画像変形と Refinement モデル，提案検出 CNN モデルによる欠損領域検出，提案補間 CNN モデルによる欠損領域補間の 3 つで構成されている．図 18 は，提案法の画像処理フローを示している．入力としては，前述した I_{ref} と I_{src} の 2 枚である．まず，従来法で用いられているオプティカルフローによる画像変形と Refinement 処理により I'_{src} を得る [19]．この時，I_{src} は I_{src} と近い明るさの画像となるように 3.2 節で述べた明るさ補正を行う．ここで，得られるオプティカルフローを用いて入力 2 枚の画像間でのオクルージョン領域 O_{src} の検出を行う．O_{src} は，オプティカルフローの従来法の基準を用いた閾値処理によって求める [62]．次に，O_{src} と I_{ref}，I_{src} を用いて提案検出用 CNN モデルで欠損領域の検出を行う．最後に，検出した欠損領域を提案補間用 CNN モデルにより補間処理を行い，欠損領域をも補間した位置合わせをおこなった画像を得られる．

提案する新たな補間 CNN モデルは，2.6.1 項で述べた畳み込みニューラルネットワークを用いた深層学習と 3.3 節で述べた Image Inpainting 手法の知見を基に多露光画像補正用に新たに提案する CNN モデルである．検出 CNN モデルは，U-Net を基にした CNN モデルであるが，従来の画像内物体領域検出とは異なりその検出の学習に直接的な教師データを用いずに学習する．また，補間 CNN モデルの構造は，2 枚の入力画像から画像の広範囲にわたる特徴と画像の詳細な特徴の 2 つを用いて補間を行うモデルを提案する．提案法では，従来手法で考慮されていなかった白とび黒つぶれと画像内物体の位置ずれの 2 つによる欠損領域について，2 つの CNN モデルを組み合わせた手法により検出と補間を行う．提案法の CNN モデルによる検出と補間により位置のあった多露光画像を得

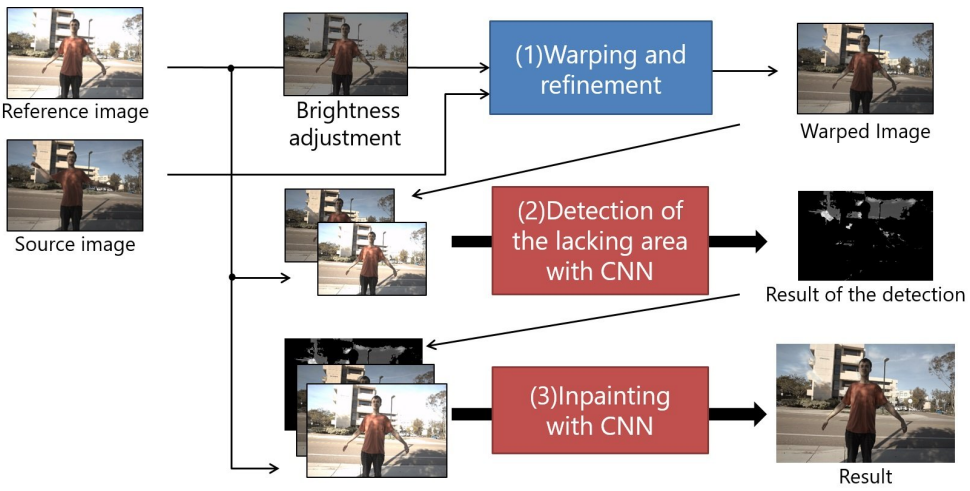


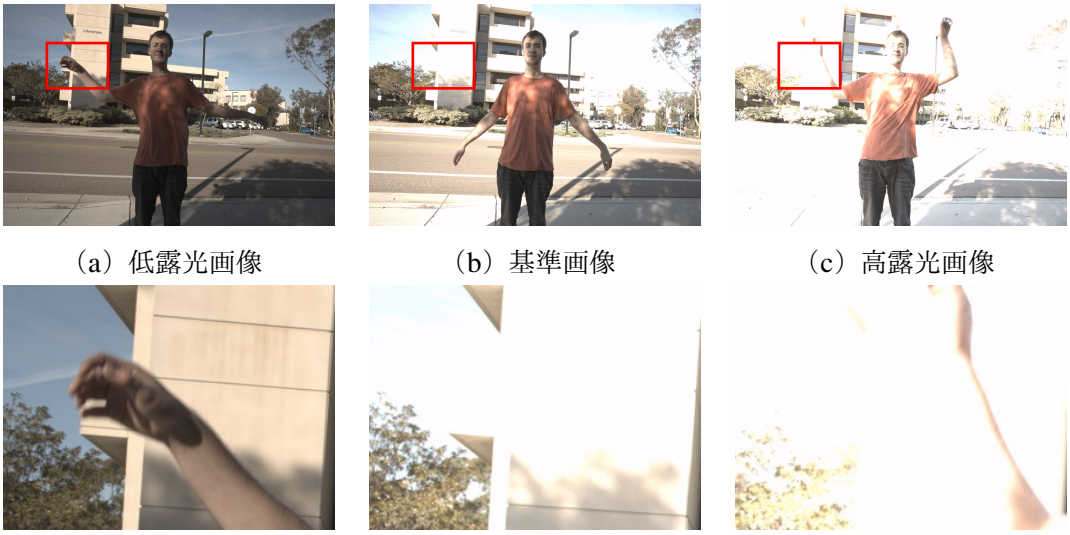
図 18　提案法の処理

られその多露光画像を合成に用いることで、結果として合成画像のアーティファクト抑制ができる．

提案する欠損領域検出 CNN モデルは，提案する 2 段階の学習により欠損領域検出を教師なし学習する．画像の特定領域を検出する CNN モデルの学習には教師あり学習が有効である．しかし，本研究で対象とするアーティファクトの原因となっている欠損領域は定義できるが，その教師データを作るのは非常に労力を要する．そこで，本研究では検出 CNN モデルと補間 CNN モデルを組み合わせた学習により検出 CNN モデルのための教師データを直接用いずに検出 CNN モデルの学習を行う．学習方法の詳細は，4.5 節で述べる．

4.2　アーティファクトを発生させる多露光画像の欠損領域の条件とその検出について

従来法の結果および抑制処理を含む最新の従来法 [17–21, 24] においてアーティファクトが発生する領域を調査したところ，アーティファクトを発生させる領域は多露光画像の画像間で情報欠損が発生している領域だとわかった．その欠損領域は，基準画像において白とびおよび黒つぶれにより画素値の情報が欠損かつその他の露光の画像においても白とび黒つぶれや画像内物体の位置ずれにより，画像の輝度値の情報が完全に欠損している領域である．図 19 に多露光画像と条件に該当する領域の拡大画像を示す．ここでは中央の中間輝度の画像が基準画像であるとする．基準画像で白とびが発生している場合は基準よりも高い露光で撮影された画像でも同様に白とびしており，反対に基準画像で黒つぶれが発生している場合は基準よりも低い露光で撮影された画像でも同様に黒つぶれが発生しているはずである．また，基準画像との位置ずれをオプティカルフローなどによって画像内物体の移動量計測が正確にでき正確に画素値を移動できたとしても，



（a）低露光画像　（b）基準画像　（c）高露光画像

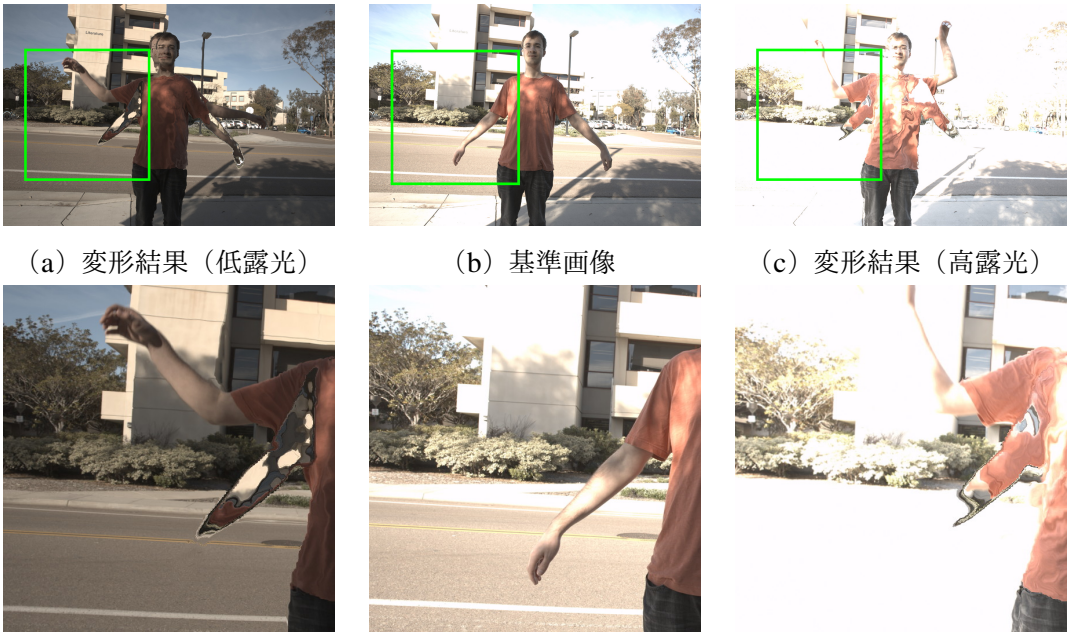
（d）拡大画像

図 19　アーティファクトが発生するの欠損領域の例

図 19 の人の手に被った領域のような場合は情報がその他の画像においても欠損したままであり，合成を行う時に手の形にアーティファクトが発生する原因となる．その領域について詳しく整理すると該当する領域は次の 2 種類の条件となる．1 つ目は図 19 のように，基準画像において白とびしているかつ，基準画像よりも低い露光を持つ画像においてその同じ位置の領域に画像内物体による位置ずれが発生している領域，2 つ目は基準画像において黒つぶれしているかつ，基準画像よりも低い露光を持つ画像においてその同じ位置の領域に画像内物体による位置ずれが発生している領域である．

この欠損領域を検出するには基準画像の白とび黒つぶれの領域の位置情報，その白とび黒つぶれと同じ領域において基準画像とその他の画像の 2 枚の間で画像内物体の位置ずれが存在する領域の情報の 2 点が必要だが，後者は位置ずれを計測することは基準画像の白とび黒つぶれにより難しく，該当する領域を閾値処理などで容易には検出できない．図 19 の（a）と（b）の画像を比較すると（a）で撮影された人の手は基準画像で大きく移動しているが，手の部分以外の領域は 2 つの画像間で位置ずれがないとわかる．しかし，従来法のオプティカルフローを用いた移動量を計測する場合，2 枚の画像間で手の姿勢が変化しておりオプティカルフローの計測が難しく，手の部分以外の領域においても基準画像の白とびにより画像特徴の一貫性などから位置ずれがないか判断するには難しい．図 20 にオプティカルフローを用いて画像変形した画像を示す．図 20（d）の拡大画像からもわかるように基準画像に完全に位置合わせした画像は得られていない．

図 21 に実際に図 19 の画像で行った従来法を用いた Refinement 処理結果と基準画像を示す．Refinement 処理で用いている画像変形はオプティカルフローを PWC-Net で計測し，画像変形を行ったものである．図 21 の（a）と（b）および（d）の各拡大画像よりオプティカルフローの計測が難しい領域において人の手のようなアーティファクトが発生



（a）変形結果（低露光）　（b）基準画像　（c）変形結果（高露光）

（d）拡大画像

図 20　オプティカルフローを用いた画像変形結果

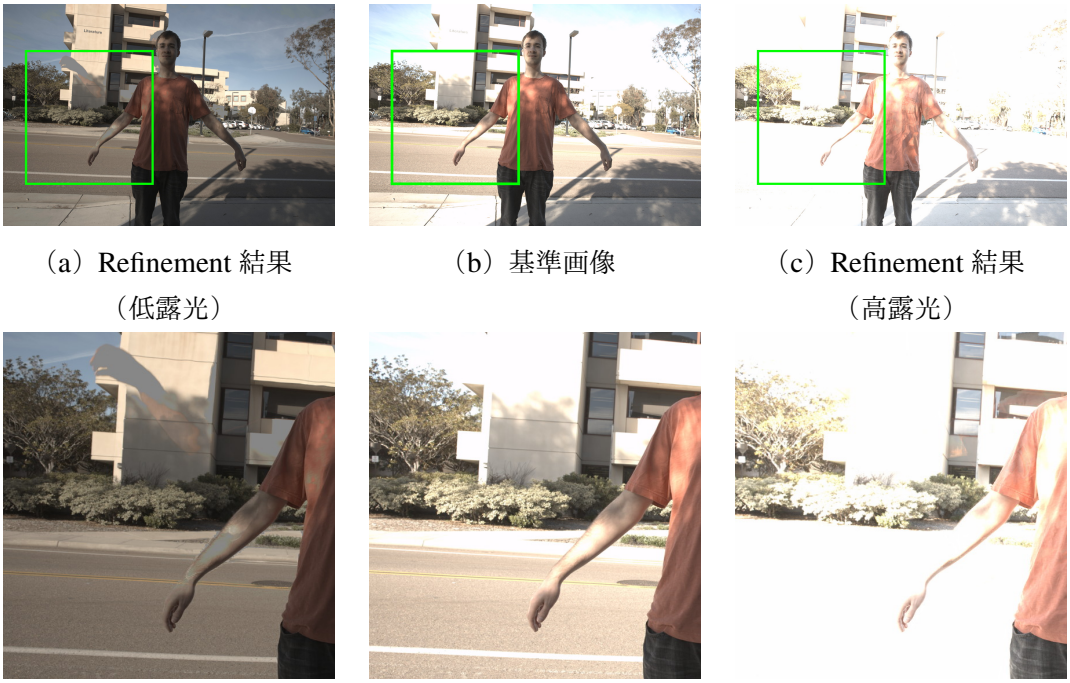
している．よって，これらの領域ではオプティカルフローおよび Refinement 処理のみではアーティファクトを発生させてしまう．

4.3　提案検出 CNN モデルを用いた動きと画素値飽和による欠損領域検出

4.2 節で述べた領域は，人が見た場合にはその位置ずれが画像全体や周囲の画像領域の情報から経験的に位置ずれがあるということはわかるため，結果としてアーティファクトが発生する欠損領域も人の目にはわかる．そこで，人の目に近い構造を持つ深層学習技術の CNN モデルであればその欠損領域の検出ができるのではないかと考え，CNN モデルを用いて欠損領域を検出する手法を提案する．提案検出 CNN モデルは，O_{src} と I_{ref}，I_{src} を入力として画像間の位置ずれおよび白とび黒つぶれによる欠損領域を検出する．O_{src} と I_{ref}，I_{src} は画像の RGB チャンネル方向に結合し，CNN モデルに入力する．O_{src} は，入力画像と同じ大きさの高さ W と幅 H を持った 2 次元の二値画像である．I_{src} において，オプティカルフローを正しく計測できないオクルージョン領域であれば O_{src} の要素は 1，その領域でなければ 0 を示す．O_{src} の d 番目の要素は文献 [62] で提案された次の条件を満たす時に 1 とする．

$$\left|f_{\mathrm{fw}}(\mathrm{d})+f_{\mathrm{bw}}(\mathrm{d}+f_{\mathrm{fw}}(\mathrm{d}))\right|^2<0.01\times\left(\left|f_{\mathrm{fw}}(\mathrm{d})\right|^2+\left|f_{\mathrm{bw}}(\mathrm{d}+f_{\mathrm{fw}}(\mathrm{d}))\right|^2\right)+0.5, \quad (67)$$

ここで，d は画素の要素番号を示す．f_{fw} は，I_{ref} を基準とした時の I_{src} のオプティカルフローであり，f_{bw} は I_{src} を基準とした時の I_{ref} である．これらのオプティカルフロー



（a）Refinement 結果（低露光）　（b）基準画像　（c）Refinement 結果（高露光）

（d）拡大画像

図 21　従来法による Refinement 結果と基準画像

は，オプティカルフローによる画像変形の時に推定されたものである．提案検出 CNN モデルを次のように式で表す．

$$W_{\mathrm{inp}} = F_{\mathrm{dtn}}(I_{\mathrm{ref}}, I'_{\mathrm{src}}, O_{\mathrm{src}}), \tag{68}$$

ここで，F_{dtn} は，提案検出 CNN モデルを示している．また，W_{inp} は，提案検出 CNN モデルの出力であり，I'_{src} の補正のための重みマップである．

提案検出 CNN モデルは，局所的な白とび黒つぶれだけでなく画像内物体の大きな動きも考慮しつつ欠損領域を検出するモデルとなるように多段のダウンサンプリングと特徴マップの結合を行う U-Net [67] を基にした構造を採用した．図 22 に提案検出 CNN モデルの構造を示す．図 22 に示すように 3 回のダウンサンプリングと特徴マップの結合を行う構造により，入力画像の複数解像度の情報を抽出する．

特に，提案モデルは U-Net とは異なりダウンサンプリングに MaxPooling ではなくストライドが 2 の畳み込み層を用いている．しかし，文献 [71] でも指摘されているように MaxPooling によるダウンサンプリングでは画像の特徴を失ってしまう場合がある．それに対して畳み込み層を用いた場合は，ダウンサンプリングで用いられるフィルタ係数が学習により獲得されるため MaxPooling に比べ必要な特徴を残しつつダウンサンプリングを行える．よって，提案検出モデルでは畳み込み層によるダウンサンプリングを用いたモデル構築を採用している．

表 1 に各層のパラメータ一覧を示す．ここで，“Output ch.” は出力特徴マップのチャンネル方向の数を表しており，“conv.” は畳み込み層を表している．提案検出 CNN モデル

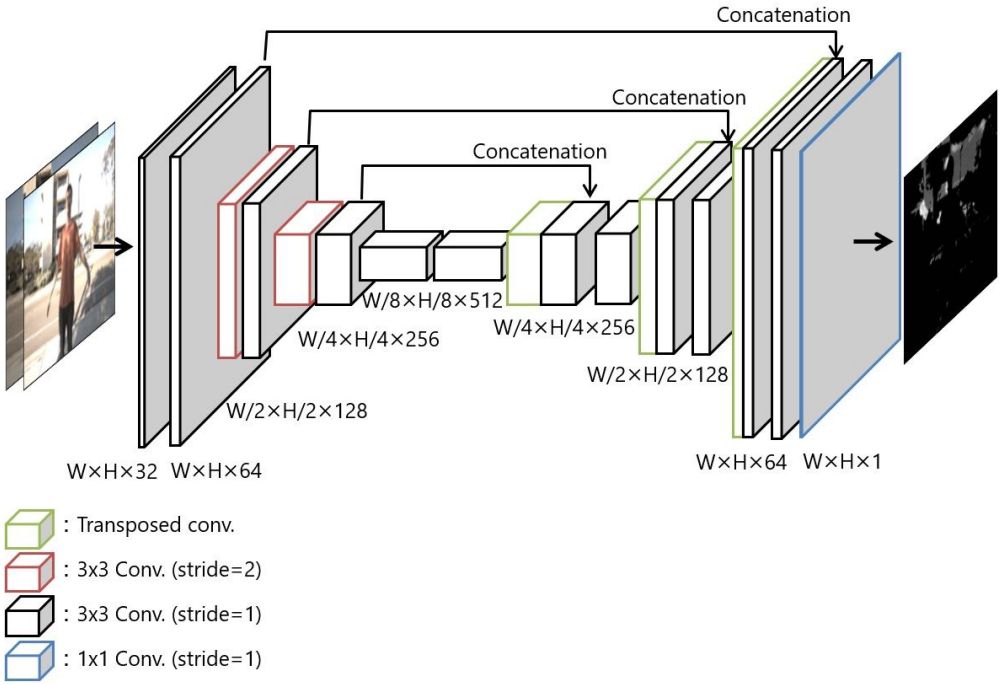


図 22　提案検出 CNN モデル

の活性化関数には，Rectified Linear Unit 関数 [36] を最後の出力層以外に用いる．出力層の後に，出力の値を $[0, 1]$ の範囲に正規化するためにシグモイド関数を適用する．これらにより，W_{inp} は $W \times H \times 1$ の大きさのテンソルとして出力される．

4.4　提案補間 CNN モデルを用いた欠損領域の補正

提案検出 CNN モデルを F_{inp} と表すと，提案補間 CNN モデルによる欠損領域の推定は以下の式で表すことができる．

$$\hat{I}_{\mathrm{src}} = F_{\mathrm{inp}}(I_{\mathrm{ref}}, I'_{\mathrm{src}}, W_{\mathrm{inp}}), \tag{69}$$

ここで，$\hat{I}_{\mathrm{src}}$ は推定した欠損領域の特徴である．また，I'_{src} は画像変形を適用した後の I_{src} 画像である．提案補間 CNN モデルの入力は，I_{ref}，I'_{src} のそれぞれの画像に W_{inp} を RGB チャンネル方向に結合した 3 次元テンソルである．CNN モデルを用いた手法において，欠損領域のマスク画像を結合した入力画像が使われるため，それを基に入力画像を定めた．最後的に，補正画像 $\hat{Y}$ は以下の式で出力される．

$$\hat{Y} = I'_{\mathrm{src}} \odot (1 - W) + \hat{I}_{\mathrm{src}} \odot W, \tag{70}$$

ここで，$\odot$ は画素ごとの積を表している．

提案補間 CNN モデルはより自然な補間を実現するため，画像の広範囲にわたる特徴と画像の詳細な特徴の両方を考慮して補間を行う構造である．図 23 にその CNN モデルの構造を示す．提案補間 CNN モデルは，（a）～（d）の 4 つの構造で構成されてい

表 1　提案検出 CNN モデルのパラメータ

Layer type	Filter size	Stride	Padding	Output ch.
Conv.	5×5	1	2	32
Conv.	3×3	1	1	64
Conv.	3×3	2	1	128
Conv.	3×3	1	1	128
Conv.	3×3	2	1	256
Conv.	3×3	1	1	256
Conv.	3×3	2	1	512
Conv.	3×3	1	1	512
Transposed conv.	4×4	2	1	256
Concatenation	-	-	-	512
Conv.	3×3	1	1	256
Transposed conv.	4×4	2	1	128
Concatenation	-	-	-	256
Conv.	3×3	1	1	128
Transposed conv.	4×4	2	1	64
Concatenation	-	-	-	128
Conv.	3×3	1	1	64
Conv.	1×1	1	0	1

るエンコーダデコーダ構造の CNN モデルである．基本的には，画像補間手法の CNN モデル [53, 57] を基にしている．画像の詳細な特徴を捉えるために Contextual attention 層 [57] を用いる．Contextual attention 層により補間に持つ物体表面の模様のような細かな特徴を抽出する．また，画像内物体の構造を表すような画像の広範囲にわたる特徴を抽出するには，通常の畳み込み層に比べ入力画像の広範囲を畳み込み演算可能にする Dilated convolution 層が有効である．よって，図 23（a）および（b）で 2 枚の画像から Contextual attention 層を組み合わせ画像のディテールを表現した特徴の抽出を行い，（c）により画像全体の特徴を捉える．（c）を通常の畳み込み層だけで構成すると受容野の広さは 53×53 であるのに対し，提案する Dilated convolution 層を組み合わせた（c）の受容野は 485×485 となりより広範囲の特徴を抽出できる．

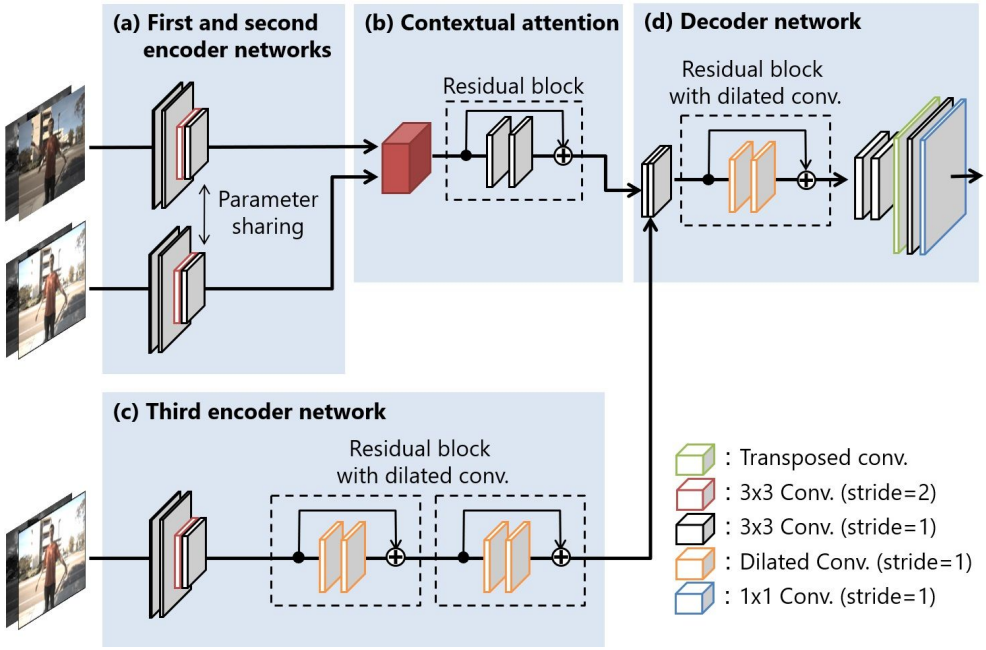


図 23　提案補間 CNN の構造

提案補間 CNN モデルの構成について入力層から順に詳細に説明する．まず初めに（a）と（c）のエンコーダネットワークで，入力画像から特徴抽出を行う．（a）は，（b）の Contextual attention 層のための特徴を行うエンコーダネットワークである．（b）の Contextual attention 層は，入力テンソルの似ている特徴を抽出するため，（a）では 2 つの入力画像で同じ特徴を抽出する必要がある．そのため，（a）ではパラメータを共通化した 2 つのエンコーダーネットワークにより特徴抽出を行う．（c）は，I_{ref} から画像の全体にわたる特徴を抽出するネットワークであり，広い領域にわたる物体位置の特徴を抽出できるように Dilated Convolution 層を用いた ResBlock で構成している．次に，（b）の Contextual Attention 層により，I'_{src} から詳細な特徴を抽出する．最後に，（d）の Decoder Network により（a）と（b）により I'_{src} から抽出された詳細な特徴と（c）により抽出された I_{ref} の広範囲にわたる特徴を合成し，入力画像と同じ大きさを持つ出力結果を得る．

表 2 にモデル各層のパラメータ一覧を示す．表 2 の “Dilation” は，2.6.3 項で示した Dilated Convolution 層のパラメータである．活性化関数としては， Leaky ReLU 関数を出力層以外のすべての層に用いる [37]．Leaky ReLU 関数は，ReLU 関数の拡張版であり負の値も許容する活性化関数である．また，出力層の後にシグモイド関数を適用し，出力の値を $[0, 1]$ に正規化する．

4.5　提案 CNN モデルの学習

本節では，提案 CNN モデルの学習について述べる．提案 CNN モデルを 2 段階の教師あり学習により，欠損領域の検出と補間を学習させる．1 段階目では提案補間 CNN モデ

表 2　提案補間 CNN モデルのパラメータ

	Layer type		Filter size	Stride	Padding	Dilation	Output ch.
(a)	Conv.		5×5	1	2	-	32
	Conv.		3×3	2	1	-	64
	Conv.		3×3	1	1	-	64
	Conv.		3×3	2	1	-	128
(b)	Contextual attention		-	-	-	-	128
	Residual block	Conv.	3×3	1	1	-	128
		Conv.	3×3	1	1	-	128
(c)	Conv.		5×5	1	2	-	32
	Conv.		3×3	2	1	-	64
	Conv.		3×3	1	1	-	64
	Conv.		3×3	2	1	-	128
	Residual block	Dilated conv.	3×3	1	2	2	128
		Dilated conv.	3×3	1	4	4	128
	Residual block	Dilated conv.	3×3	1	8	8	128
		Dilated conv.	3×3	1	16	16	128
(d)	Concatination		-	-	-	-	256
	Residual block	Dilated conv.	3×3	1	2	2	256
		Dilated conv.	3×3	1	4	4	256
	Transposed conv.		4×4	2	1	-	128
	Conv.		3×3	1	1	-	128
	Transposed conv.		4×4	2	1	-	64
	Conv.		3×3	1	1	-	32
	Conv.		1×1	1	1	-	3

ルの Image Inpainting の事前学習を行い，2 段階目に多露光画像のデータセットを用いた 2 つの提案 CNN モデルの学習を行う．提案検出 CNN モデルの事前学習については，検出 CNN モデルのタスクに近いデータセットが存在しないため行わない．この学習で用いるデータセットの画像については，その画素値を $[0, 1]$ の範囲に正規化したものを使用する．

まず 1 段階目の学習について述べる．1 段階目は，Image Inpainting のデータセットを用いた提案補間 CNN モデルの事前学習である．入力画像 I として画像のランダムな場

所を欠損させたものを下記の式で用意する．

$$I = x \odot m, \tag{71}$$

ここで，x はデータセットにある加工前の画像であり，m は 0 か 1 の値を持つ二値マスク画像であり欠損領域の場所を示している．x は，この事前学習の正解画像としても用いる．m は，画像のランダムな場所を正方形に欠損させるように作成する．現実シーンでは欠損が現れる場所は画像の端から端まで一定ではないため，一様分布を持つランダムな値によって欠損領域の座標を決めた m を用いる．事前学習の出力結果 $\hat{x}$ は，提案補間 CNN モデル F_{inp} を用いて次の式で表わされる．

$$\hat{x} = I \odot (1 - m) + F_{\mathrm{inp}}(I, I, m) \odot m, \tag{72}$$

ここで，I は I_{ref} および I'_{sec} の代わりに 2 回入力するため，2 回記述している．また，この事前学習では損失関数として以下の式で表わされる画素ごとの Mean Absolute Error（MAE）を用いる．

$$L_{\mathrm{pre}}(\hat{x}, x) = \frac{1}{3wh} \sum_{i=1}^{w} \sum_{j=1}^{h} \|\hat{\boldsymbol{x}}_{i,j} - \boldsymbol{x}_{i,j}\|_1, \tag{73}$$

ここで，w と h は入力画像の高さと幅を表している．また $\hat{\boldsymbol{x}}_{i,j}$ および $\boldsymbol{x}_{i,j}$ は，それぞれ $\hat{x}$ と x の画像の (i, j) 画素における RGB の画素値を持つベクトルである．この損失関数の最適化には ADADELTA 法を用いたミニバッチ確率的勾配降下法を用いる [44].

次に，2 段階目の多露光画像のデータセットを用いた 2 つの提案 CNN モデルの学習である．提案 CNN モデルの学習では，入力する補正対象の画像 I_{src} の様々な露光に対応させる学習をさせる．そのため，入力する補正対象の画像 I_{src} は露光の高い画像と低い画像の両方からランダムに選択し学習に用いる．式 (70) より，提案法では提案検出 CNN モデルと提案補間 CNN モデルの結果を微分可能な演算で用いている．そのため，検出 CNN モデルと補間 CNN モデルの計算グラフを接続できるため，2 つの CNN モデルに対して同時に誤差逆伝播法による学習が可能である．よって，2 段階目の学習では 1 つの損失関数により 2 つの CNN モデルを同時に学習させる．提案補間 CNN モデルには，1 段階目の事前学習で学習した各層の重みパラメータを初期値として用いる．2 段階目の損失関数は，1 段階目と同じように以下の式で表わされる MAE を用いる．

$$L_{\mathrm{main}}(\hat{Y}, Y) = \frac{1}{3wh} \sum_{i=1}^{w} \sum_{j=1}^{h} \|\hat{\boldsymbol{Y}}_{i,j} - \boldsymbol{Y}_{i,j}\|_1, \tag{74}$$

ここで，$\hat{\boldsymbol{Y}}_{i,j}$ および $\boldsymbol{Y}_{i,j}$ は，それぞれ $\hat{Y}$ と Y の画像の (i, j) 画素における RGB の画素値を持つベクトルである．この損失関数の最適化には，1 段階目と同じく ADADELTA 法を用いたミニバッチ確率的勾配降下法を用いる．この学習で，提案補間 CNN モデルを GAN の Generator ネットワークとして新たに Discriminator ネットワークを追加して敵対的学習により学習することも考慮したが，学習が安定せず学習ができなかった．

4.6　4 章のまとめ

本章は本研究で提案する CNN モデルを用いた多露光画像の位置合わせ補正手法についてその詳細を述べた．提案法は位置ずれと白とび黒つぶれによる欠損領域を提案検出 CNN モデルによって検出し提案補間 CNN モデルによって画像情報の補間を行いアーティファクトの発生が抑制された多露光画像を作成する．提案 CNN モデルは 2 段階の学習を行い検出と補間を学習する．

第 5 章

提案手法と従来法を用いたアーティファクト抑制効果の実験

5.1 概要

本章では，4 章の提案法によって作成した多露光画像と欠損領域の検出および従来の合成手法との比較実験について述べる．

真値が備わっているデータセットを用いた定量的な比較および実際のシーンを撮影した画像を用いた視覚的な比較により，提案法のアーティファクト抑制性能について検証を行う．また，提案法を多露光画像合成手法の前処理として用いた場合のアーティファクト抑制効果について検証するため比較実験を行う．まず 5.2 節において本研究で行った実験の条件を示す．特に，提案法の学習条件や用いるデータセットについて述べる．

本実験では，4 つの比較実験を行い提案法の補正性能について評価および考察を行う．まず，5.3 節で提案法により出力される位置合わせ多露光画像の真値画像との比較について述べる．次に，5.4 節で従来法のアーティファクト抑制処理を含む多露光画像合成手法との合成画像での比較について述べる．また，5.5 節で従来法のアーティファクト抑制を含む HDR 画像合成手法との比較について述べる．さらに，5.6 節で従来法の多露光画像合成法の前処理として提案法を用いた場合の実験について述べる．

5.2 実験条件

5.2.1 データセット

提案法の学習には，それぞれの段階で 2 種類のデータセットを用いた．第 1 段階の提案補間 CNN モデルの事前学習では画像補間手法の学習でよく用いられる Place2 データセットを用いた [72]．Place2 データセットの学習用画像から 1 万枚をランダムに選び，さらに学習用と Validation 用にそれぞれ 9000 枚と 1000 枚に分割して用いた．それらの画像を 256×256 の大きさに縮小し，4.5 節で述べたように画像内のランダムな場所に

128×128 の大きさの欠損領域を作ったものを入力画像として用い加工前の画像を正解画像として用いた．第 2 段階の学習では Kalantari らの HDR 画像データセットを用いた．そのデータセットには 74 セットの位置ずれを含む多露光画像とそれと同一シーンの位置ずれがない HDR 画像が含まれている．学習では多露光画像のうち中間露光の画像を基準画像として用いた．正解画像は HDR 画像からカメラレスポンス関数を用いて生成した位置ずれなし多露光画像を作成し用いた．学習用画像はその画像サイズを 1500×1000 から 384×256 の大きさへ縮小しまた 256×256 の大きさで画像を 64pixel ずつずらしながら切り取り，さらに水平方向反転と 90 度ずつの回転を加えるデータ拡張を行った．最終的に，それらの処理によって学習に用いる基準画像と補正対象画像および正解画像のデータセットは合計で 4440 セットとなった．

評価に用いる画像データは，Kalantari らのテスト用データセットと Karaduzovic-Hadziabdic らのデータセット，Tursun らのデータセットを用いる [17, 73, 74]．Kalantari らのデータセットには入力用多露光画像とアーティファクトのない正解用 HDR 画像が含まれている．そのため，正解画像が必要な定量的評価指標の計算には Kalantari らのデータセットを用いる．正解用 HDR 画像は，入力用多露光画像と同じシーンで撮影した動きのない多露光画像から Debevec らの手法 [4] を用いて HDR 画像合成を行ったものである．また，入力多露光画像の中で中間輝度画像が基準画像であり，正解 HDR 画像と同じ画像内物体の位置になっている．本章では，正解の HDR 画像を表示するために Photomatix Pro 5.1 のトーンマッピング処理によりダイナミックレンジを圧縮して表示し視覚的評価を行う [75]．PhotomatixPro5.1 のトーンマッピング処理では，デフォルト設定かつノイズ除去やその他の除去設定は全て無効にした設定で HDR 画像からダイナミックレンジを圧縮した画像を生成した．

本章の各表や図などで表記する Kalantari らのテスト用データセットの画像と定量的評価の表における画像名の対応を図 24 に示す．

5.2.2　従来法と提案法の実装方法および学習条件

提案 CNN モデルの実装および学習には Python の深層学習ライブラリである Chainer v7 と CUDA v10.0 および CUDA の深層学習用ライブラリである CUDNN v7 を用いた．学習に用いた計算機は CPU に Intel 製 Xeon E5-2620 v4 と GPU に Nvidia 社製 Geforce GTX1080ti を搭載している．第 1 段階の学習ではバッチサイズ 12 で 225000 イタレーションの学習を行い，第 2 段階ではバッチサイズ 10 で 177600 イタレーション行った．前述した計算機で第 1 段階の学習に約 48 時間，第 2 段階の学習に約 38 時間かかった．

本実験では，公開されている従来法のプログラムを用いて各手法の結果を得て，視覚的評価および定量的評価を行っている．深層学習を用いている従来法は全てコードおよび学習済み CNN モデルが公開されているため，それを用いた．それらを各手法の著者らが用いているのと同じバージョンの Python ライブラリを用いて動作させ，コードも入力画

図 24　Kalantari らのテスト用データセットの入力多露光画像（a）および正解多露光画像（b）とトーンマッピング済み正解 HDR 画像（c）の一覧

像の読み込み部分の変更など最低限の変更にとどめ結果を出力した．Python ライブラリはオープンソースものでありそのバージョンによって実装が変わり出力結果に影響する場合があるため，指定されたバージョンのものを用いた．また，定量的評価指標は Peak Signal-to-Noise Ratio（PSNR），Multi-scale structual similarity（MS-SSIM），High Dynamic Range Visual Difference Predictor 2（HDR-VDP2）を用い，PSNR と MS-SSIM については MATLAB の Image Processing Toolbox に実装されているものを使用し，HDR-VDP2 については公開されている MATLAB コードを用いた．

表 3　提案法補正画像の PSNR 評価結果（低露光画像）

	PSNR [dB]			
	入力	画像変形	Refinement	提案法補正
Image 1	17.19	16.09	20.76	24.27
Image 2	15.78	15.41	18.70	26.56
Image 3	18.28	18.68	20.42	26.15
Image 4	25.28	28.11	30.63	35.02
Image 5	19.90	27.04	30.07	34.73
Image 6	18.60	23.62	25.40	25.80
Image 7	22.86	24.07	24.96	27.03
Image 8	24.45	24.82	30.47	32.87
Image 9	16.85	22.05	26.13	27.95
Image 10	29.81	30.25	26.08	23.3
Image 11	25.39	27.96	24.96	22.9
Image 12	26.33	28.75	29.95	32.6
Image 13	19.17	23.24	26.23	27.8
Image 14	16.98	22.76	24.70	26.2
Image 15	17.85	18.66	18.94	20.8
Average on 15 sets	20.98	23.43	25.23	27.62

5.3　提案法による多露光画像と正解多露光画像との比較

本節では，提案法によって作成した多露光画像と真値の HDR 画像から作成した位置ずれのない正解多露光画像との比較を行う．HDR 画像から多露光画像を生成するには，HDR 画像が輝度値に対してリニアな値を持つため任意のカメラレスポンス関数を適用する必要がある．カメラレスポンス関数により，画素値が異なるため特に PSNR の評価指標に影響する．データセットにおいて入力画像を撮影したカメラのカメラレスポンス関数を得られれば良いが，そのカメラレスポンス関数は公開されていなかったため本実験では使用できなかった．そのため本実験の PSNR は参考値であり，位置ずれの補正およびアーティファクトの抑制では構造的類似度が近ければアーティファクトの原因となる正解画像において輝度値が変わってしまうため PSNR の値は参考値である．本比較実験は Multi-scale SSIM と PSNR を用いた定量的評価と視覚的評価により提案法の結果画像の評価を行った．正解画像を作成する時に文献 [76] でモデル化された Agfacolor Future 100DC のカメラレスポンス関数を用いた．

表 3 および表 4 に正解画像と各段階の結果画像での PSNR スコアを，表 5 および表 6

表 4　提案法補正画像の定量的評価結果（高露光画像）

	PSNR [dB]			
	入力	画像変形	Refinement	提案法補正
Image 1	18.33	19.81	24.09	27.02
Image 2	15.16	15.69	23.78	27.03
Image 3	18.59	20.36	25.78	28.09
Image 4	20.34	22.51	24.52	29.14
Image 5	13.97	22.19	26.28	29.35
Image 6	13.04	17.43	20.17	21.12
Image 7	14.33	14.46	14.50	16.05
Image 8	16.48	16.93	25.52	25.60
Image 9	11.21	16.07	20.94	22.36
Image 10	23.58	22.78	17.35	17.4
Image 11	20.56	21.59	16.09	16.1
Image 12	21.22	24.09	24.86	27.3
Image 13	12.88	17.69	21.16	22.3
Image 14	12.75	19.35	21.96	23.6
Image 15	17.84	19.32	19.83	21.0
Average on 15 sets	16.68	19.35	21.79	23.59

に MS-SSIM のスコアを示す．各表のスコアは，正解画像に対して入力画像や画像変形処理結果，Refinement 処理結果，また，図 25 および図 26，図 27，図 28，図 29 にそれぞれの入力画像と基準画像，オプティカルフローによる画像変形結果，Refinement 処理結果，提案法補正結果，正解画像を示す．MS-SSIM の値は $[0, 1]$ の範囲であり，数値が高いほど正解画像に近くより良い画像と言える．

表 3，表 4，表 5 および表 6 から，オプティカルフローによる画像変形，Refinement 処理および提案法補正処理により一部を除き PSNR および MS-SSIM の値が上昇していることがわかる．Refinement 処理と提案法補正結果の値を比べると，その平均値および各画像全ておいて評価値が上昇しており，提案法により構造的に類似した画像を作成できていることがわかる．また，各画像の MS-SSIM も 0.9 を超えており，正解画像に対して構造的に高い類似度を持つ画像であると評価できる．特に，図 25（e）と（f）を比べると建物の壁面にある人の手の影のようなアーティファクトを抑制できていることがわかる．さらに，図 26（e）と（f）から提案法補正結果では建物の壁面の細かいテクスチャを復元できていることがわかる．よって，提案法は大きな動きによるアーティファクトおよび画像の詳細な特徴を復元する補正ができることがわかる．

表 3 および表 5 の Image 1 と Image 2 においては，オプティカルフローによる画像変形

表 5　提案法補正画像の MS-SSIM 評価結果（低露光画像）

	MS-SSIM（範囲 [0,1]）			
	入力	画像変形	Refinement	提案法補正
Image 1	0.798	0.835	0.900	0.920
Image 2	0.796	0.780	0.909	0.942
Image 3	0.833	0.878	0.912	0.923
Image 4	0.836	0.951	0.968	0.982
Image 5	0.662	0.929	0.969	0.984
Image 6	0.491	0.831	0.917	0.929
Image 7	0.927	0.953	0.965	0.965
Image 8	0.907	0.922	0.960	0.968
Image 9	0.441	0.812	0.929	0.942
Image 10	0.962	0.966	0.956	0.944
Image 11	0.917	0.950	0.948	0.934
Image 12	0.934	0.963	0.973	0.980
Image 13	0.530	0.854	0.931	0.942
Image 14	0.539	0.884	0.931	0.948
Image 15	0.838	0.888	0.873	0.891
Average on 15 sets	0.761	0.893	0.936	0.946

の PSNR と MS-SSIM の値が下がっているがこれはオプティカルフローの誤検出を原因とする画像変形の誤りによるものだと考えられる．図 25（d）と（g）の画像から，Image 1 のオプティカルフローによる画像変形はその誤検出や未検出により画像内物体の位置合わせを全て正確に行えてはいないことがわかる．特に画像内の人物の胸あたりに背景の建物から画素値を移動させてしまっており画素値が大きく変化しているため PSNR の値が低くなったと考えられる．図 25（h）と（k）でも同様に誤った画像変形が行われているが PSNR の値は上昇してはいるものの，他の画像や平均値の上昇幅と比べると低く，この画像変形の誤りが強く影響していると考えられる．また，Image2 の低露光画像においても背景の建物が大きく変形してしまっている．この領域においては構造的にも大きく変形しているため，この画像では PSNR と MS-SSIM の値が低くなっていると考えられるこれも同様に，オプティカルフローの誤検出により背景の建物が大きく変形してしまっているためだと考えられる．

図 25 の低露光画像において、提案法により画像情報を補間することで背景の建物にある影の様なアーティファクトの削減が行えている．図 25（d）と（e）を比べると画像変形によって正しく位置ずれを補正できていない領域を Refinement 処理によりある程度補正していることがわかる．しかし，図 25（g）と（e）を比べると，人の腕の形に影のよ

表 6　提案法補正画像の定量的評価結果（高露光画像）

	MS-SSIM（範囲 [0,1]）			
	入力	画像変形	Refinement	提案法補正
Image 1	0.762	0.841	0.915	0.935
Image 2	0.784	0.802	0.946	0.955
Image 3	0.799	0.856	0.952	0.962
Image 4	0.717	0.909	0.942	0.952
Image 5	0.478	0.907	0.954	0.964
Image 6	0.247	0.748	0.870	0.882
Image 7	0.839	0.863	0.907	0.930
Image 8	0.860	0.879	0.942	0.944
Image 9	0.183	0.715	0.885	0.898
Image 10	0.926	0.935	0.888	0.891
Image 11	0.904	0.924	0.884	0.887
Image 12	0.872	0.919	0.945	0.955
Image 13	0.248	0.782	0.888	0.898
Image 14	0.386	0.836	0.912	0.922
Image 15	0.832	0.884	0.890	0.902
Average on 15 sets	0.656	0.853	0.915	0.925

うなアーティファクトを発生させていることがわかる．ここで図 25（a）と（b）および（e），（g）を比べると，Refinement 処理で影のようなアーティファクトを発生させていた領域は（a）および（b）にある位置ずれかつ基準画像における白とびによる欠損領域であることがわかる．図 25（g）と（f），（e）を比べると，（f）の提案法補正結果では（g）にあった影のようなアーティファクトを削減し（g）の正解画像に近い画像を作成できていることがわかる．図 25 でも同様に白い車の部分に発生していた手の影のようなアーティファクトを抑制し，さらに背景にある建物の壁面のテクスチャを復元できている．

しかし，図 27 の低露光画像の提案法補間結果では補間が十分にできていない場合が発生している．この領域は，他の画像と比較すると大きく欠損している領域である．このような，大きな欠損領域の補間は Image Inpainting の分野においても非常に難しいタスクである．また，学習に用いた Kalantari らデータセットにはこのような非常に大きい欠損領域を含む画像は少ない．そのため，提案法は大きい欠損領域に対して未だ自然な補間が行われていないと考えられる．当問題に対しては画像補間手法の最新手法の知見を用いて CNN モデルの構造を構築し，大きな欠損領域のある画像を数多く有するデータセットを用いて学習することで解決できるのではないかと考えられる．


（a）補正対象画像（低露光）


（b）基準画像


（c）補正対象画像（高露光）


（d）画像変形結果


（e）Refinement 処理結果


（f）提案法補正結果（低露光）


（g）正解画像（低露光）


（h）画像変形結果


（i）Refinement 処理結果


（j）提案法補正結果（高露光）


（k）正解画像（高露光）

図 25　Image 1 の提案法補正結果

5.4　多露光画像合成における定量的評価および視覚的評価による比較

本節では，アーティファクト抑制を含む従来の合成手法と提案法の合成結果画像における定量評価的および視覚的な比較実験について述べる．


（a）補正対象画像（低露光）


（b）基準画像


（c）補正対象画像（高露光）


（d）画像変形結果


（e）Refinement 処理結果


（f）提案法補正結果（低露光）


（g）正解画像（低露光）


（h）画像変形結果


（i）Refinement 処理結果


（j）提案法補正結果（高露光）


（k）正解画像（高露光）

図 26　Image 2 の提案法補正結果

5.4.1　Kalantari らのデータセットを用いた定量的評価および視覚的評価による比較

本実験では提案法の定量評価を Kalantari らのデータセットを用いて行う [17]．比較する従来法として多露光画像合成手法の最新手法である Ma らの手法と比較を行

（a）補正対象画像（低露光）　（b）基準画像　（c）補正対象画像（高露光）

（d）画像変形結果　（e）Refinement 処理結果

（f）提案法補正結果（低露光）　（g）正解画像（低露光）

（h）画像変形結果　（i）Refinement 処理結果

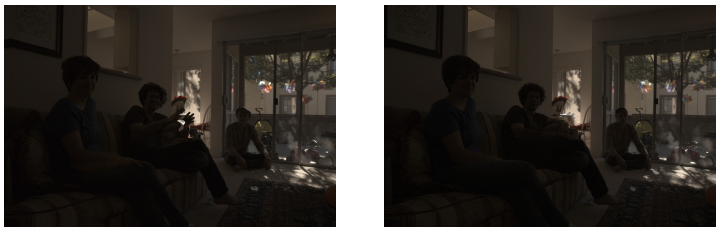
（j）提案法補正結果（高露光）　（k）正解画像（高露光）

図 27　Image 3 の提案法補正結果

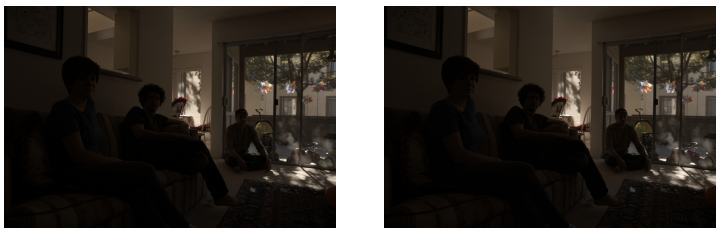
う [19, 24]．Kalantari らのデータセットには評価用データとして自然なシーンで撮影された位置ずれを含む多露光画像と位置ずれのない真値 HDR 画像のデータが 15 セット含まれている．Kalantari らのデータセットの真値用 HDR 画像は位置ずれのない多露光画像から Debevec らの手法 [4] を用いて合成した HDR 画像である [17]．

また合成画像の比較では，画像評価指標として一般的に用いられる PSNR および

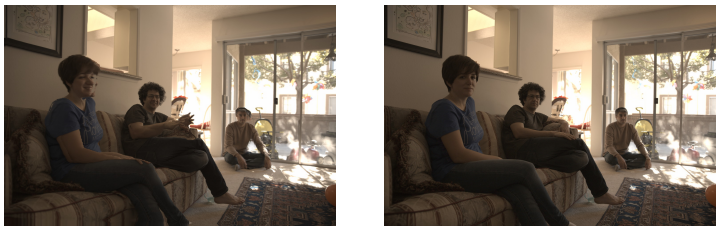
（a）補正対象画像（低露光）　（b）基準画像　（c）補正対象画像（高露光）



（d）画像変形結果　（e）Refinement 処理結果



（f）提案法補正結果（低露光）　（g）正解画像（低露光）



（h）画像変形結果　（i）Refinement 処理結果

（j）提案法補正結果（高露光）　（k）正解画像（高露光）

図 28　Image 4 の提案法補正結果

MS-SSIM を用いる．PSNR は画像の評価において真値画像とその輝度値がどれだけ近いか評価する指標である．そこで，多露光画像合成の手法とも比較するため本実験において HDR 画像から入力多露光画像と同じ枚数と露光を持つ多露光画像を作成し，その多露光画像を Mertens らの手法 [8] で多露光画像合成したものを GT として用いた．HDR 画像から多露光画像を作成する時に文献 [76] でモデル化された Agfacolor Future 100DC の

（a）補正対象画像（低露光）　（b）基準画像　（c）補正対象画像（高露光）

（d）画像変形結果　（e）Refinement 処理結果

（f）提案法補正結果（低露光）　（g）正解画像（低露光）

（h）画像変形結果　（i）Refinement 処理結果

（j）提案法補正結果（高露光）　（k）正解画像（高露光）

図 29　Image 5 の提案法補正結果

カメラレスポンス関数を用いた．

表 7 より，提案法により従来の多露光画像合成よりも定量的に高いか同じ評価の画像を生成できているとわかる．表 7 の PSNR より，すべての画像において従来法よりも良い値を示している．また，表 7 の MS-SSIM からは，Image 1 から Image 5 で良い評価値でありより構造的に真値画像と近くアーティファクトを抑制していることがわかる．平

（a） 入力画像（基準画像：中間露光画像）

（b） [24]　　（c）提案法 + [8]　　（d）正解画像

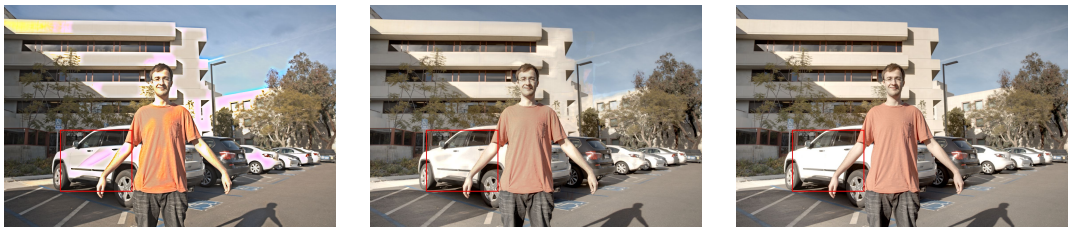
図 30　Image 1 の多露光画像を用いたアーティファクト抑制を含む多露光画像合成手法の結果

均値においては，従来法よりも 0.005 低い値でありほぼ同等の評価といえる．

また，図 30，図 31，図 32，図 33，図 34 から，視覚的にも従来法に比べアーティファクトを抑制した合成画像を得られている．Ma らの手法の結果画像では，各画像において入力画像にはないピンク色に近い色のアーティファクトが発生している．このピンク色のアーティファクトは，Ma らの手法で合成画像の生成処理で用いられている RGB 値ごとのヒストグラムマッチング処理により発生している．図 30，図 31，図 32，図 33，図 34 の各図の（a）と（b）を見ると入力画像において画素値の飽和している領域にピンク色の不自然な色が出てしまっていることがわかる．よって，この画素値の飽和した領域において，今回のテスト画像では R の成分が強く出てしまいアーティファクトが発生していると考えられる．

それに対して，図 30，図 31，図 32，図 33，図 34 の（b）と（c）の画像を比べると，提案法はそれらのアーティファクトを抑制した結果を出力できているとわかる．特に，画像内物体の位置ずれが大きい人の腕に出ているアーティファクトを抑制できている．しかし提案法は，図 32（d）と（e）の画像とそれらの拡大画像から多露光画像からわかるとおり，位置ずれと白とび黒つぶれによる欠損領域が大きい場合にその欠損領域の情報を自然に推定して補間することができていないことがわかる．5.3 節の図 27（f）においても同様に補間しきれていない領域である．これは 5.3 節で述べた大きな欠損領域の補

（a） 入力画像（基準画像：中間露光画像）



（b） [24]　（c）提案法 + [8]　（d）正解画像

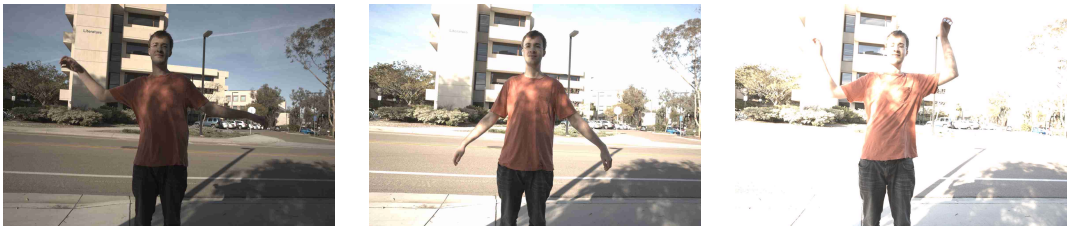
図 31　Image 2 の多露光画像を用いたアーティファクト抑制を含む多露光画像合成手法の結果

間しきれていないことによる影響である．

5.4.2　自然なシーンの画像を用いた視覚的評価による比較

本項では実際のシーンで取得した多露光画像を用いて合成結果におけるアーティファクト抑制効果の評価実験について述べる．本実験ではアーティファクト抑制の評価によく用いられる Karaduzovic-Hadziabdic らのデータセットと Tursun らのデータセットを用いた [73, 74]．しかし，それらの実際のシーンの画像データセットでは真値となる動きのない多露光画像および合成後の画像はないため定量評価指標を用いた比較は行えないので視覚的評価による比較のみ行う．

図 35 と図 36 に Karaduzovic-Hadziabdic らのデータセットと Tursun らのデータセットでの合成結果画像を示す．図 35 と図 36 において（a）は入力多露光画像，（b）-（d）は各手法での合成結果である．図 35 の一番下段の画像は結果画像の拡大画像である．Karaduzovic-Hadziabdic らのデータセットは複雑な動きをしているものを含む画像が多く，図 35（b）では滑り台に影のようなアーティファクトが発生している．さらに図 35（c）では滑り台の一部が欠損しているように見えるアーティファクトが発生している．しかし，図 35（d）の提案法の結果画像ではそれらの領域においてアーティファクトが



（a） 入力画像（基準画像：中間露光画像）

（b） [24]　　（c）提案法 + [8]　　（d）正解画像

図 32　Image 3 の多露光画像を用いたアーティファクト抑制を含む多露光画像合成手法の結果

発生しておらず従来法に比べ高いアーティファクト抑制効果を有していることがわかる．また，図 36（b）-（d）の中段および下段の画像は上段の画像の拡大画像である．中段は結果画像において赤枠の領域の拡大画像であり下段は緑枠の領域の拡大画像である．図 36（b）および（c）の拡大画像から従来法の結果画像にはアーティファクトが発生してしまっていることがわかる．しかし，図 36（d）の提案法の結果画像ではそれらのアーティファクトを抑制できており自然な画像を作成できているとわかる．よってこれらの結果から提案法は最新手法に比べ視覚的にアーティファクトを抑制できていることがわかる．

5.5　HDR 画像合成手法における定量的評価および視覚的評価による比較

本節では，アーティファクト抑制処理を含む HDR 画像合成手法と提案法の結果を用いて HDR 画像合成をする場合との比較実験について述べる．HDR 画像の定量的評価指標として HDR-VDP2 を用いる [2]．HDR-VDP2 は，真値画像を用いて HDR 画像を評価する手法であり，HDR 画像の評価に広く用いられている指標である．提案法は多露光画像の補正手法であるため，HDR 画像合成のためにアーティファクト抑制のない Debevec らの手法と PhotomatixPro5.1 の HDR 画像合成機能を用いる [75]．本実験では単純な HDR 画像合成法として用いるため，PhotomatixPro5.1 のノイズ除去等の追加機能はすべ

（a） 入力画像（基準画像：中間露光画像）

（b） [24]　　（c）提案法 + [8]　　（d）正解画像

図 33　Image 4 の多露光画像を用いたアーティファクト抑制を含む多露光画像合成手法の結果

て無効化して用いる．従来法としては，Niu らの HDR 画像合成手法を用いる [21].

表 8 に従来法および提案法の出力画像の HDR-VDP2 の値を示す．また図 37 から図 41 において（a）と（b）-（e）に Image 1 と Image 5 の入力多露光画像と各手法の HDR 画像合成結果および真値画像を示す．図 37 から図 41 の各 HDR 画像は，表示のために PhotomatixPro5.1 のトーンマッピング処理によりダイナミックレンジ圧縮をかけたものである．トーンマッピング処理では，PhotomatixPro5.1 のデフォルト設定を使用しノイズ除去などは用いていない．

表 8 の結果より，HDR-VDP2 の結果では 2021 年の最新手法である Niu らの手法に数値的には劣っているということがわかった．Niu らの結果と提案法＋ Debevec らの手法を比べると Debevec らの手法を用いた結果は非常に低い数値となっている．この結果は，Debevec らの手法では正確な HDR 画像を合成するために露光時間が必要だが，Kalantari らのデータセットには相対露光の値しかなく，本実験では MATLAB に実装されている相対露光による HDR 画像合成を用いて合成しているためだと考えられる．PhotomatixPro5.1 を合成処理に用いた場合は，Niu らの手法に近い結果を出せており，HDR-VDP2 の値は合成手法の精度に大きく依存していることがわかる．

図 37 から図 41 の各手法の合成結果を比較すると提案法を用いた合成結果は，Niu らの手法の結果と同等か少し劣る結果となっている．しかし提案法は，Niu らの手法に比べ任意の枚数を持つ多露光画像に適用できる点と HDR 合成とアーティファクト抑制の問

（a） 入力画像（基準画像：中間露光画像）

（b） [24]　　（c）提案法 + [8]　　（d）正解画像

図 34　Image 5 の多露光画像を用いたアーティファクト抑制を含む多露光画像合成手法の結果

題を分ける点において学術的貢献がある．図 37 から図 38 において Niu らの手法と提案法補正結果を用いた HDR 合成結果を比較すると，人の動きによるアーティファクトの抑制に関しては大きな違いがなくどちらも抑制できていることがわかる．図 37 の提案法結果では，画像右上にある空の部分などに真値画像にはない雲のようなアーティファクトが発生してしまっている．しかし，図 37（b）や図 39（b）の Niu らの手法の結果では建物と空の境界において不自然なボケが発生してしまっているが，図 37（d）や図 39（d）の提案法結果では発生していない．

提案法の HDR-VDP2 や視覚的評価においては，従来法よりも良い部分と劣る部分が存在するが，提案法にはその原理的に従来法よりも優れている部分があると考えられる．Niu らの手法と提案法を用いた合成を比較すると，Niu らの手法では入力する多露光画像は 3 枚固定であり提案法は原理的には任意の枚数を補正可能である．また，Niu らの手法では多露光画像を入力として HDR 画像を直接推定する手法であり，提案法のように合成とアーティファクト抑制処理を分離できず Mertens らの手法のような多露光画像合成にアーティファクト抑制を応用できない．これらの点において，提案法は最新手法にくらべ研究と応用への発展がしやすいという学術的貢献がある．提案法と Niu らの手法の処理や学習について比較すると，Niu らの手法では GAN を用いた学習を取り入れている．GAN は画像補間の細かいテクスチャの復元に寄与する学習方法であるが，その学習

表 7　多露光画像合成の各手法結果画像の PSNR および MS-SSIM

	PSNR [dB]		MS-SSIM	
	[24]	提案法 + [8]	[24]	提案法 + [8]
Image 1	21.40	**26.88**	0.920	**0.956**
Image 2	20.53	**27.86**	0.933	**0.957**
Image 3	22.07	**27.56**	0.900	**0.942**
Image 4	25.96	**31.14**	0.955	**0.963**
Image 5	26.03	**32.07**	0.951	**0.973**
Image 6	23.65	**23.85**	**0.969**	0.914
Image 7	22.80	**29.58**	**0.962**	0.959
Image 8	22.25	**28.11**	**0.967**	0.957
Image 9	23.38	**24.95**	**0.962**	0.919
Image 10	21.70	**28.54**	0.962	**0.964**
Image 11	23.55	**26.34**	**0.968**	0.945
Image 12	22.92	**29.64**	**0.961**	**0.961**
Image 13	24.32	**24.96**	**0.966**	0.919
Image 14	20.85	**26.39**	0.957	**0.951**
Image 15	21.65	**22.22**	**0.944**	0.931
15 セットの平均	22.87	**27.34**	**0.952**	0.947

を安定して行うことは難しい．提案法でも GAN を用いた学習について取り組んだが，その学習が安定しなかったため本研究の提案法に組み込むことはできなかった．学習が安定する条件を見つけることができれば，提案法に GAN を組み込むことで Niu らの手法と同等以上の品質の画像が得られると考えられる．

5.6　提案法を前処理として用いた場合の比較実験

本節では提案法を従来法の多露光画像合成手法の前処理として用いた場合のアーティファクト抑制効果について比較実験について述べる．比較に用いる合成手法は Liu および Wang の手法と Li らの手法および Ma らの手法である [13, 15, 24]．これらの手法は代表的な合成手法でありアーティファクト抑制を考慮した手法も含まれている．本実験には Kalantari らのデータセットと Karaduzovic-Hadziabdic らのデータセットを用いる [17, 73].

図 42 と図 43 に Kalantari らのデータセットと Karaduzovic-Hadziabdic らのデータセットの画像に対する合成結果を示す．図 42 と図 43 において（a）と（b）-（g）はそれぞれ入力画像と各手法のみの結果とそれら手法の前処理として提案法を組み込んだ場合

（a） 入力画像（基準画像：中間露光画像）

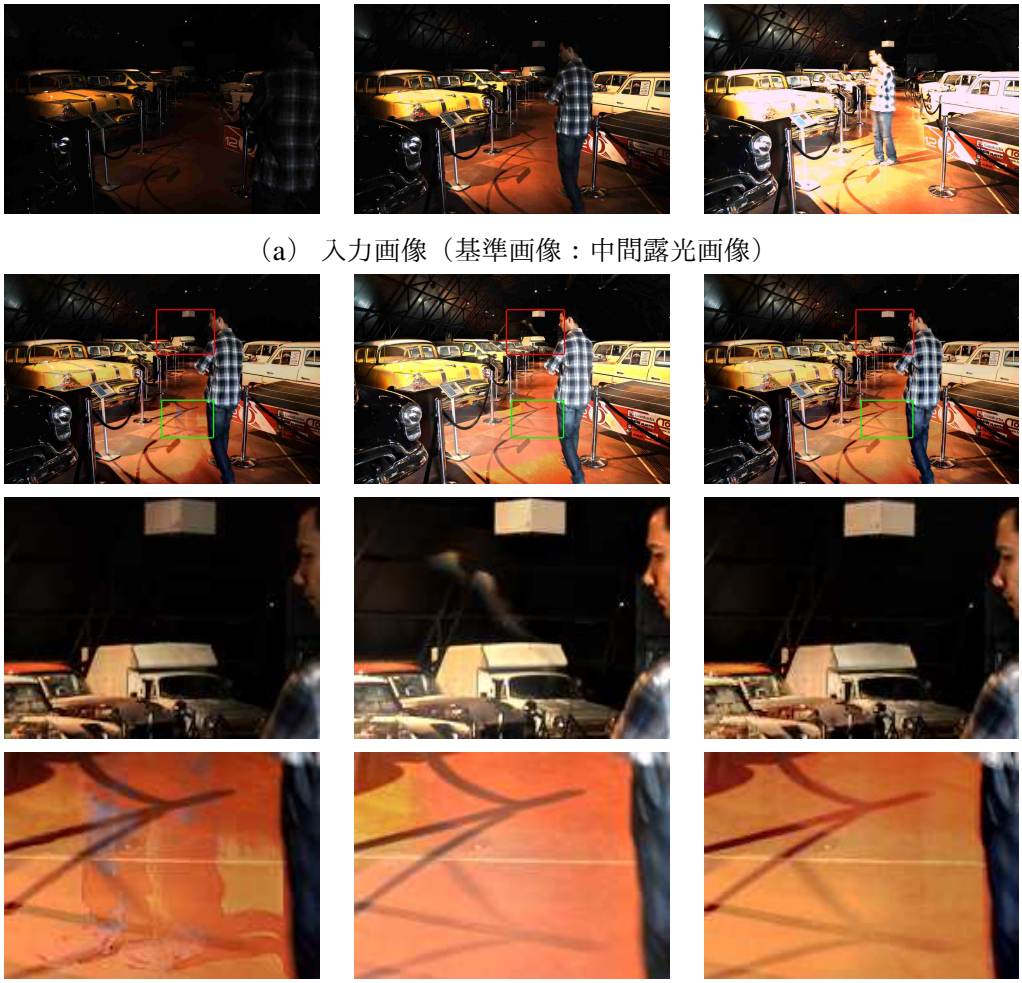
（b） [19]　　（c） [24]　　（d） 提案法 + [8]

図 35　Karaduzovic-Hadziabdic らのデータセットを用いたアーティファクト抑制を含む多露光画像合成手法の結果画像

の結果画像を示す．各図の（b）-（g）の下段の画像は各結果画像の拡大画像である．図 42（b）と（d），（f）の画像では建物に透明な手のようなアーティファクトが発生していることがわかる．そのアーティファクトは図 42（c）と（e）および（g）では抑制されていることがわかる．また，図 43（b）と（d）の画像では半透明の人影に見えるアーティファクトが発生しており図 43（f）の画像では床の一部に床の模様とは異なるアーティファクトが発生している．図 43（c）と（e），（g）では図 42 のと同様にアーティファクトが抑制されていることがわかる．これらの結果より，提案法は多露光画像合成手法のアーティファクト抑制のための前処理として効果的であるといえる．

5.7　まとめ

本章では，本研究の提案法による画像補正による効果と提案法を用いた合成とアーティファクト抑制を含む従来法の合成後の結果画像においてアーティファクト抑制効果の比較実験を行った．3 つのデータセットを用い，PSNR，MS-SSIM，HDR-VDP2 を用いた定量評価と視覚的評価を行った．提案法による補正により，定量的評価の値が上がり視覚的な劣化を補正できていることが結果からわかり，提案法には効果があることを示した．また，多露光画像合成の従来法に比べ提案法を含む合成では，定量的評価および視覚的に同等かよりよいアーティファクト抑制性能があることを示した．さらに，従来の HDR 画像合成手法と提案法との結果比較により，提案法は HDR-VDP2 の評価では劣る



（a） 入力画像（基準画像：中間露光画像）

（b） [19]　　（c） [24]　　（d） 提案法 + [8]

図 36　Tursun らのデータセットを用いたアーティファクト抑制を含む多露光画像合成手法の結果画像

が，視覚的には従来法と同等かより優れた抑制効果を有することを示した．加えて，従来法の前処理として提案法を用いた場合も，従来法の抑制処理の有無にかかわらずアーティファクトを抑制する効果があることを比較実験により示した．

表 8　Kalantari らのデータセットにおける HDR 画像合成の各手法結果画像の HDR-VDP2

	HDR-VDP2		
	[21]	提案法 + [4]	提案法 + Photomatix Pro 5.1
Image 1	**75.47**	57.42	67.22
Image 2	**78.95**	57.48	71.20
Image 3	**76.07**	59.54	71.49
Image 4	**79.36**	50.14	76.73
Image 5	**82.00**	48.99	77.96
Image 6	**73.82**	45.51	65.00
Image 7	**78.69**	46.85	71.05
Image 8	**85.27**	47.26	70.43
Image 9	**71.57**	44.43	69.78
Image 10	**82.76**	46.82	67.73
Image 11	**80.90**	44.37	62.37
Image 12	**78.52**	46.82	70.97
Image 13	**70.60**	44.47	69.42
Image 14	**74.73**	45.53	69.38
Image 15	**67.15**	50.18	56.64
15 セットの平均	**77.06**	49.06	69.16

（a） 入力多露光画像（基準画像：中間露光画像）

（b） [21]

（c） 提案法 + [4]（相対露光）

（d）提案法 + Photomatix Pro 5.1

（e） 正解画像

図 37　Image 1 の多露光画像を用いたアーティファクト抑制を含む HDR 画像合成手法の結果

（a） 入力多露光画像（基準画像：中間露光画像）

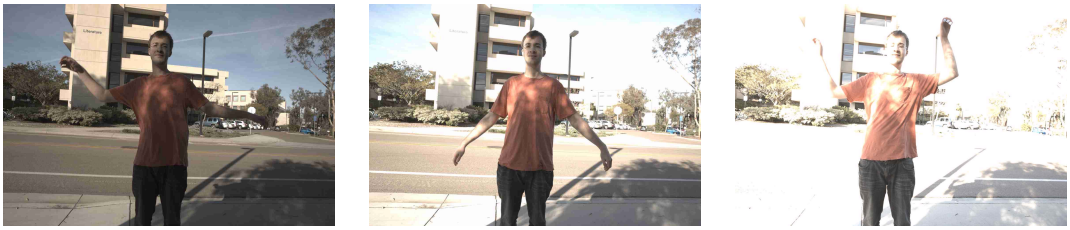
（b） [21]

（c） 提案法 + [4]（相対露光）

（d）提案法 + Photomatix Pro 5.1

（e） 正解画像

図 38　Image 2 の多露光画像を用いたアーティファクト抑制を含む HDR 画像合成手法の結果



（a） 入力多露光画像（基準画像：中間露光画像）

（b） [21]

（c） 提案法 + [4]（相対露光）

（d）提案法 + Photomatix Pro 5.1

（e） 正解画像

図 39　Image 3 の多露光画像を用いたアーティファクト抑制を含む HDR 画像合成手法の結果

（a） 入力多露光画像（基準画像：中間露光画像）

（b） [21]

（c） 提案法 + [4]（相対露光）

（d）提案法 + Photomatix Pro 5.1

（e） 正解画像

図 40　Image 4 の多露光画像を用いたアーティファクト抑制を含む HDR 画像合成手法の結果

（a） 入力多露光画像（基準画像：中間露光画像）

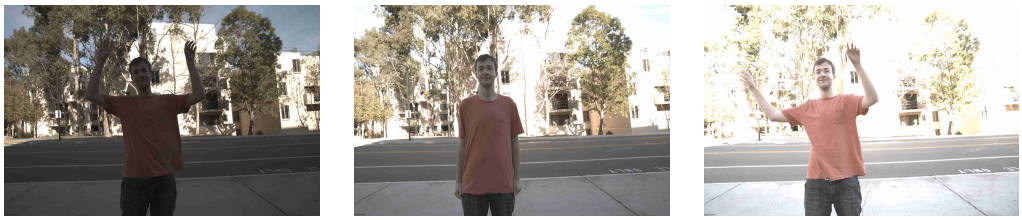
（b） [21]

（c） 提案法 + [4]（相対露光）

（d）提案法 + Photomatix Pro 5.1

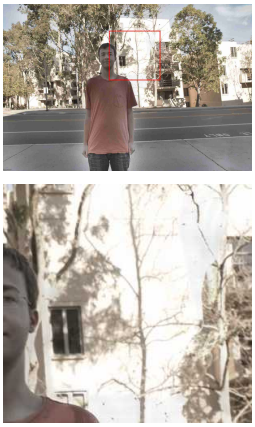
（e） 正解画像

図 41　Image 5 の多露光画像を用いたアーティファクト抑制を含む HDR 画像合成手法の結果



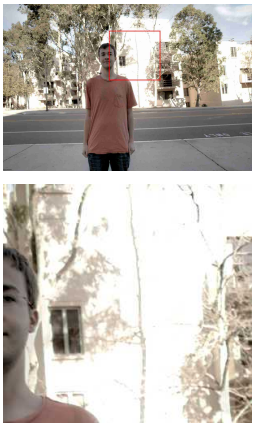
（a） 入力画像（基準画像：中間露光画像）

（b） [13]

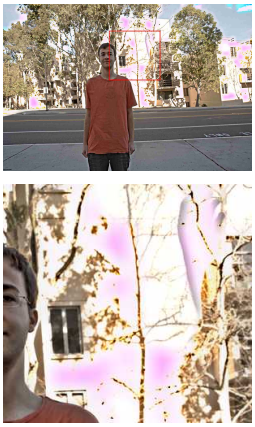


（c）提案法 + [13]

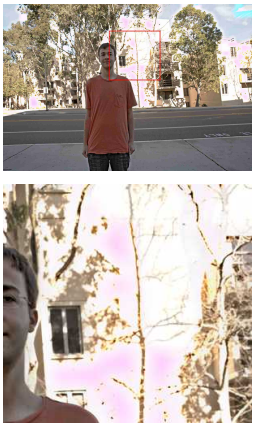
（d） [15]



（e） 提案法 + [15]

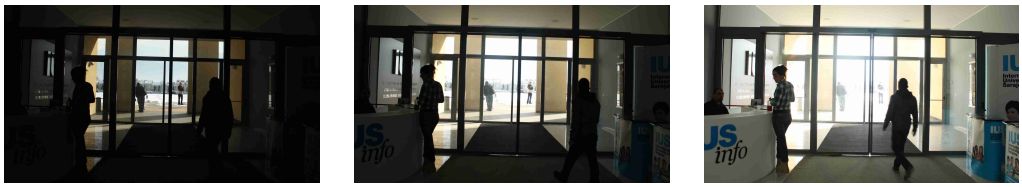


（f） [24]



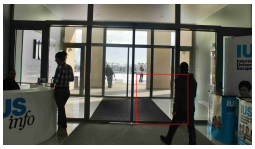
（g） 提案法 + [24]

図 42　Kalantari らのデータセットを用いた従来法とそれら従来法の前処理として提案法用いた場合の合成結果画像



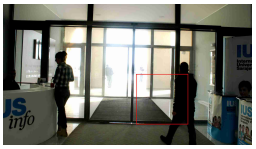
（a） 入力画像（基準画像：中間露光画像）

（b） [13]

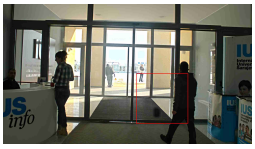


（c） 提案法 + [13]

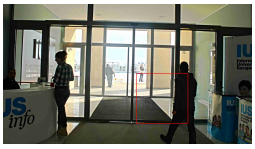
（d） [15]



（e） 提案法 + [15]



（f） [24]



（g） 提案法 + [24]

図 43　Karaduzovic-Hadziabdic らのデータセットを用いた従来法とそれら従来法の前処理として提案法用いた場合の合成結果画像

第 6 章

結論

本研究では，多露光画像合成および HDR 画像合成において大きな問題となるアーティファクトの抑制を実現することを目的として，深層学習を用いた多露光画像の位置合わせ手法を提案し従来法に比べ高いか同等の抑制性能を持ち，多露光画像を用いる応用技術の前処理としても有効な手法の提案を達成できた．

1 章にて，デジタルカメラについての特性や多露光画像合成や HDR 画像合成についてその歴史と，本研究で着目したアーティファクト抑制の手法についてその手法や問題といった本研究の背景について述べた後，本研究の目的および成果，本論文の構成について述べた．

2 章にて，デジタルカメラや多露光画像の合成，機械学習や深層学習など本研究を理解する上で必要な基礎理論について述べた．

3 章にて，アーティファクト抑制を含む多露光画像合成手法や HDR 画像合成手法，画像補間手法など本研究に関連する研究について述べた後，それらの研究に対する本研究の立ち位置について述べた．

4 章にて，従来法においてアーティファクトを発生させる多露光画像の画像間で発生する欠損領域について定義し，提案する多露光画像の補正手法について述べたアーティファクトの原因となっている欠損領域は，基準画像の白とび黒つぶれとそれ以外の露光画像における白とび黒つぶれおよび物体の位置ずれによる欠損が同時に発生している多露光画像間の領域である．さらに，従来法のオプティカルフロー計測による画像変形および CNN モデルを用いた Refinement 処理に加えて，アーティファクトを発生させる欠損領域を深層学習を用いて検出し補間する多露光画像補正手法を提案した．従来のアーティファクト抑制を含む合成手法ではオプティカルフローによる画像内物体の位置合わせと CNN モデルで推定した重みによる重み付き合成によって位置ずれを補正し合成している．それに対して提案法は，つの提案 CNN モデルを用いてその領域を検出し補間することでアーティファクトの発生を抑制した位置合わせ多露光画像を生成する手法を実現している．これにより，合成処理にはアーティファクト抑制処理を含まない手法でも提案法を前処理として適用することでアーティファクト抑制が可能になり，従来よりもアーティファクトを抑制した合成画像を作成できる．また，欠損領域の検出および補間

を行う CNN モデルについてその構造と学習法について提案した．補間を行う CNN モデルは，画像補間手法で用いられている CNN モデルの技術を基に 2 枚の画像を入力とし，欠損領域補間のため入力画像の広い領域にわたる特徴と細かい領域にあるテクスチャなどの特徴の 2 つを抽出し補間するモデルとした．さらに，検出と補間の CNN モデルの学習には 2 段階の学習を用いることにより検出 CNN モデルの教師データを用意せずに学習を行う．

5 章にて，提案法の補正によるアーティファクト抑制の効果を確かめるため，4 つの実験について述べた．1 つ目の実験として，提案法補間結果について真値画像との定量的および視覚的な比較を行い，提案法によってより真値画像に近く位置ずれを補正した多露光画像を生成できていることを示した．2 つ目の実験として，アーティファクト抑制を含む HDR 画像合成の従来法と提案法の HDR 画像におけるその抑制効果を定量的および視覚的に比較評価し，実際のシーンで撮影された多露光画像において提案法が従来法に比べ高いアーティファクト抑制効果を有していることを示した．3 つ目の実験として，アーティファクト抑制を含む多露光画像合成の従来法と提案法の合成画像におけるその抑制効果を定量的および視覚的に比較する評価を行った．さらなる実験として，従来の多露光画像を用いた合成手法の前処理として提案法を適用した場合と適用しない場合を定量的および視覚的に評価し比較した．その実験を通して，提案法によりアーティファクト抑制処理を含む含まないに関わらず従来法のアーティファクト抑制を行えることを示した．しかし，これらの実験より提案法には定義した欠損領域が広い場合自然な補間を行えないという制限があることがわかった．広い欠損領域の画像情報の補間は画像補間の分野でも難しい問題である．

更なる提案法の発展としては，広い欠損領域にも対応するために最新の画像補間手法の知見を用いて提案法の補間 CNN モデルを構築し，その広い欠損領域を多く含む多露光画像データセットを用いて学習することが考えられる．現在では，人の視覚特性に近いダイナミックレンジを持ったイメージセンサも研究開発されているが一般的に普及するには至っておらず，一般的に普及している LDR しか取得できないデジタルカメラを用いて HDR 画像や白とび黒つぶれのない画像を得るためには多露光画像を用いた合成技術が不可欠である．さらに新しく開発された高いダイナミックレンジを持つイメージセンサと合成技術を組み合わせることで人の視覚特性を超えたセンシングや自動化システムが開発できる．よって，多露光画像を用いたアーティファクト抑制や HDR 画像合成，多露光画像合成技術の研究および手法の提案には意義がある．本研究でもアーティファクト問題の解決法を提案しておりその後の発展の調査および得られた知見を基にして研究中である．本研究の成果により多露光画像を用いた画像合成および合成後の画像を用いたコンピュータビジョンの分野がさらに発展していくことを願う．

謝辞

本研究は著者が電気通信大学大学院情報理工学研究科博士後期課程に在籍中に行ったものであり，本研究を行うにあたり多くの方の御協力と御指導，御支援を賜りました．長岡技術科学大学の修士課程から公私ともにご指導とご助言，ご支援を厚く賜った主任指導教員および本論文の審査委員会の主査である電気通信大学 情報理工学研究科 吉田太一助教に心より感謝し厚く御礼申し上げます．

また，関連論文や研究内容についてご指導ご助言を賜りました指導教員および論文審査委員会の委員である電気通信大学 情報理工学研究科 張 熙 教授に深く感謝し厚く御礼申し上げます．ご多忙の中論文審査委員会の委員を引き受けてくださいました，電気通信大学　情報理工学研究科　野村 英之 教授，高橋 弘太　准教授，劉 志 准教授に深く感謝申し上げます．

また，本研究の関連論文においてご助言ご指導を賜りました長岡技術科学大学大学院電気電子情報工学専攻 岩橋 政宏教授に深く感謝申し上げます。さらに，本研究についてご相談させていただきご助言を賜りました慶應義塾大学 理工学部 池原 雅章教授にこの場を借りて深く感謝いたします．

また，博士後期課程在学中に JST 次世代研究者挑戦的研究プログラムの電気通信大学独自ネットワーク形成を行う開発主導型博士学生研究・教育支援プログラムおよび電気通信大学大学院 大学院博士後期課程奨学金の支援をいただきました深く感謝いたします．

最後に，博士後期課程に進学した私を支えてくださいました祖父と両親，友人たちに心から感謝いたします。

研究業績一覧

関連論文

査読付き論文誌論文

(1) Isana funahashi，Taichi Yoshida，Zhang Xi and Masahiro Iwahashi，“Image Adjustment for Multi-exposure Images Based on Convolutional Neural Networks,” IEICE Transactions on Information and Systems，Vol.E105-D，No.1，pp.123–133，Jan. 2022.（2021 年 9 月採録決定済）

その他研究業績

査読付き論文誌論文

(1) Yo Umeki, Isana Funahashi, Taichi Yoshida and Masahiro Iwahashi, "Salient Object Detection With Importance Degree," in IEEE Access, vol.8, pp.147059-147069, 2020.

国際学会（査読あり）

(1) Isana funahashi，Naoki Yamashita, Taichi Yoshida and Masaaki Ikehara，“High Reflection Removal Using CNN with Detection and Estimation,” Asia Pacific Signal and Information Processing Association Annual Summit and Conference 2021 accepted, oral presentation, Dec. 2021.

国内学会（査読なし）

(1) 山下尚樹, 船橋勇那, 吉田太一, “強反射領域の検出を用いた反射除去手法,” in Proc. 電子情報通信学会 総合大会 2021.

(2) 東海林佳昭, 小笠原志朗, 柴田朋子, 船橋勇那，“業務実施者間の相互支援に向けた画像・音声からの相補的業務経験把握手法,” 電子情報通信学会ソサイエティ大会講演論文集, B-14-6, Sep. 2021.

参考文献

[1] Z. Wang, E. Simoncelli, and A. Bovik, "Multiscale structural similarity for image quality assessment," in *Proc. the Thrity-Seventh Asilomar Conference on Signals, Systems Computers*, vol. 2, 2003, pp. 1398–1402.

[2] R. Mantiuk, K. J. Kim, A. G. Rempel, and W. Heidrich, "Hdr-vdp-2: A calibrated visual metric for visibility and quality predictions in all luminance conditions," in *Proc. of ACM Special Interest Group on Computer Graphics and Interactive Techniques Conf.*, no. 40, 2011, pp. 1–14.

[3] A. Tomaszewska and R. Mantiuk, "Image Registration for Multi-exposure High Dynamic Range Image Acquisition," in *Proc. the 15-th Intl. Conf. in Central Europe on Comput. Graphics, Visualization and Comput. Vis.*, 2007, pp. 49–56.

[4] P. E. Debevec and J. Malik, "Recovering high dynamic range radiance maps from photographs," in *Proc. the 24th Annual Conf. Comput. Graphics and Interactive Techniques*, 1997, pp. 369–378.

[5] G. Ward, "Fast, Robust Image Registration for Compositing High Dynamic Range Photographs from Hand-Held Exposures," *Journal of Graphics Tools*, vol. 8, no. 2, pp. 17–30, 2003.

[6] O. Gallo, N. Gelfandz, W.-C. Chen, M. Tico, and K. Pulli, "Artifact-free High Dynamic Range imaging," in *Proc. IEEE Intl. Conf. Computational Photography*, 2009, pp. 1–7.

[7] S. Raman and S. Chaudhuri, "Bilateral filter based compositing for variable exposure photography," in *Proc. Eurographics*, 2009, pp. 1–4.

[8] T. Mertens, J. Kautz, and F. V. Reeth, "Exposure fusion: A simple and practical alternative to high dynamic range photography," *Computer Graphics Forum*, vol. 28, no. 1, pp. 161–171, 2009.

[9] R. Shen, I. Cheng, J. Shi, and A. Basu, "Generalized random walks for fusion of multi-exposure images," *IEEE Trans. Image Process.*, vol. 20, no. 12, pp. 3634–3646, Dec 2011.

[10] P. Sen, N. K. Kalantari, M. Yaesoubi, S. Darabi, D. B. Goldman, and E. Shechtman, "Robust patch-based hdr reconstruction of dynamic scenes," *ACM Trans. Graphics*, vol. 31, no. 6, pp. 1–11, 2012.

[11] C. Lee, Y. Li, and V. Monga, "Ghost-Free High Dynamic Range Imaging via Rank Minimization," *IEEE Signal Process. Letters*, vol. 21, no. 9, pp. 1045–1049, 2014.

[12] T. Sakai, D. Kimura, T. Yoshida, and M. Iwahashi, "Hybrid method for multi-exposure image fusion based on weighted mean and sparse representation," in *Proc. European Signal Process. Conf.*, 2015, pp. 809–813.

[13] Y. Liu and Z. Wang, "Dense SIFT for ghost-free multi-exposure fusion," *J. Vis. Commun. Image Represent.*, vol. 31, pp. 208–224, 2015.

[14] K. R. Prabhakar and R. V. Babu, "Ghosting-free multi-exposure image fusion in gradient domain," in *Proc. IEEE Intl. Conf. Acoustics, Speech and Signal Process.*, 2016, pp. 1766–1770.

[15] Z. Li, Z. Wei, C. Wen, and J. Zheng, "Detail-Enhanced Multi-Scale Exposure Fusion," *IEEE Trans. Image Process.*, vol. 26, no. 3, pp. 1243–1252, 2017.

[16] Z. Wang, Q. Liu, and T. Ikenaga, "Robust Ghost-Free High-Dynamic-Range Imaging by Visual Salience Based Bilateral Motion Detection and Stack Extension Based Exposure Fusion," *IEICE Trans. Fundamentals of Electronics, Communications and Comput. Sciences*, vol. E100.A, no. 11, pp. 2266–2274, 2017.

[17] N. K. Kalantari and R. Ramamoorthi, "Deep high dynamic range imaging of dynamic scenes," *ACM Trans. Graphics*, vol. 36, no. 4, 2017.

[18] S. Wu, J. Xu, Y.-W. Tai, and C.-K. Tang, "Deep high dynamic range imaging with large foreground motions," in *Proc. the European Conf. Comput. Vis.*, 2018, pp. 117–132.

[19] K. R. Prabhakar, R. Arora, A. Swaminathan, K. P. Singh, and R. V. Babu, "A Fast, Scalable, and Reliable Deghosting Method for Extreme Exposure Fusion," in *Proc. IEEE Intl. Conf. Computational Photography*, 2019, pp. 1–8.

[20] Q. Yan, D. Gong, Q. Shi, A. van den Hengel, C. Shen, I. Reid, and Y. Zhang, "Attention-guided network for ghost-free high dynamic range imaging," in *Proc. 2019 IEEE/CVF Conf. Comput. Vis. Patt. Recogni.*, 2019, pp. 1751–1760.

[21] Y. Niu, J. Wu, W. Liu, W. Guo, and R. W. H. Lau, "Hdr-gan: Hdr image reconstruction from multi-exposed ldr images with large motions," *IEEE Trans. Image Process.*, vol. 30, pp. 3885–3896, 2021.

[22] T. Kartalov, Z. Ivanovski, and L. Panovski, "A real time global motion compensation for multi-exposure imaging algorithms," in *Proc. IEEE EUROCON - Intl. Conf. Comput. as a Tool*, 2011, pp. 1–4.

[23] J. Hu, O. Gallo, K. Pulli, and X. Sun, "HDR Deghosting: How to Deal with Saturation?" in *Proc. IEEE Conf. Comput. Vis. Patt. Recognit.*, 2013, pp. 1163–1170.

[24] K. Ma, H. Li, H. Yong, Z. Wang, D. Meng, and L. Zhang, "Robust Multi-Exposure Image Fusion: A Structural Patch Decomposition Approach," *IEEE Trans. Image Process.*, vol. 26, no. 5, pp. 2519–2532, 2017.

[25] M. Okuda, "High dynamic range image coding," *The Journal of The Institute of Image*

Information and Television Engineers, vol. 64, no. 3, pp. 299–305, 2010.

[26] 株式会社ニコン, “高速、高解像で、暗いところから明るいところまでを瞬時に撮像できる積層型 cmos イメージセンサーを開発,” 2021, https://www.nikon.co.jp/news/2021/0217_cmos_01.htm (2021/10/19 参照).

[27] R. Szeliski, コンピュータービジョンアルゴリズムと応用. 共立出版, 2013, 玉玉木徹，福嶋慶繁，飯山将晃，鳥居秋彦，栗田多喜夫，波部斉，林昌希，野田雅文訳.

[28] R. Erik, H. Wolfgang, D. Paul, P. Sumanta, W. Greg, and M. Karol, *High Dynamic Range Imaging 2nd Edition Acquisition, Display, and Image-Based Lighting*. Elsevier Inc., 2010.

[29] M. Ryo, “私のブックマーク：高ダイナミックレンジ画像処理,” 人工知能, vol. 36, no. 1, pp. 90–96, 2021.

[30] J. M. Ogden, E. H. Adelson, J. R. Bergen, and P. J. Burt, “Pyramid-based computer graphics,” *RCA Engineer*, vol. 5, no. 30, 1985.

[31] G. Ian, B. Yoshua, and C. Aaron, *Deep Learning*. The MIT Press, 2016.

[32] Y. LeCun and I. Misra, “Self-supervised learning: The dark matter of intelligence,” 2021, https://ai.facebook.com/blog/self-supervised-learning-the-dark-matter-of-intelligence/ (2021/10/25 参照).

[33] M. Kevin, P., *Machine Learning : A Probabilistic Perspective*. The MIT Press, 2012.

[34] A. Tavanaei, M. Ghodrati, S. R. Kheradpisheh, T. Masquelier, and A. Maida, “Deep learning in spiking neural networks,” *Neural Networks*, vol. 111, pp. 47–63, 2019.

[35] Y. Bengio, P. Simard, and P. Frasconi, “Learning long-term dependencies with gradient descent is difficult,” *IEEE Trans. Neural Networks*, vol. 5, no. 2, pp. 157–166, 1994.

[36] V. Nair and G. E. Hinton, “Rectified linear units improve restricted boltzmann machines,” in *Proc. Intl. Conf. Machine Learning*, 2010, pp. 807–814.

[37] A. L. Maas, A. Y. Hannun, and A. Y. Ng, “Rectifier nonlinearities improve neural network acoustic models,” in *Proc. Intl. Conf. Machine Learning*, 2013.

[38] K. Fukushima, “Neocognitron: A self-organizing neural network model for a mechanism of patt. recognit. unaffected by shift in position,” *Biological Cybernetics*, vol. 36, no. 4, pp. 193–202, 1980.

[39] A. Krizhevsky, I. Sutskever, and G. E. Hinton, “Imagenet classification with deep convolutional neural networks,” in *Proc. Intl. Conf. Neural Info. Process. Systems*, 2012, pp. 1097–1105.

[40] Google, “Cloud tpu,” 2021, https://cloud.google.com/tpu (2021/10/25 参照).

[41] A. Paszke, S. Gross, F. Massa, A. Lerer, J. Bradbury, G. Chanan, T. Killeen, Z. Lin, N. Gimelshein, L. Antiga, A. Desmaison, A. Kopf, E. Yang, Z. DeVito, M. Raison, A. Tejani, S. Chilamkurthy, B. Steiner, L. Fang, J. Bai, and S. Chintala, “Pytorch: An imperative style, high-performance deep learning library,” in *Advances in Neural Information Processing Systems*, vol. 32, 2019.

[42] B. Polyak, “Some methods of speeding up the convergence of iteration methods,” *USSR Computational Mathematics and Mathematical Physics*, vol. 4, no. 5, pp. 1–17, 1964.

[43] J. Duchi, E. Hazan, and Y. Singer, “Adaptive subgradient methods for online learning and stochastic optimization,” *Journal of Machine Learning Research*, vol. 12, no. 61, pp. 2121–2159, 2011.

[44] M. D. Zeiler, “ADADELTA: an adaptive learning rate method,” *arXiv:1212.5701*, 2012.

[45] Y. LeCun, B. Boser, J. S. Denker, D. Henderson, R. E. Howard, W. Hubbard, and L. D. Jackel, “Backpropagation applied to handwritten zip code recognition,” *Neural Computation*, vol. 1, no. 4, pp. 541–551, 1989.

[46] M. Lin, Q. Chen, and S. Yan, “Network in network,” in *Proc. Intl. Conf. Learn. Represent.*, 2014.

[47] J. Long, E. Shelhamer, and T. Darrell, “Fully convolutional networks for semantic segmentation,” in *Proc. IEEE Conf. Comput. Vis. Patt. Recognit.*, 2015, pp. 3431–3440.

[48] C. Szegedy, W. Liu, Y. Jia, P. Sermanet, S. Reed, D. Anguelov, D. Erhan, V. Vanhoucke, and A. Rabinovich, “Going deeper with convolutions,” in *Proc. IEEE Conf. Comput. Vis. Patt. Recognit.*, 2015.

[49] K. He, X. Zhang, S. Ren, and J. Sun, “Deep residual learning for image recognition,” in *Proc. IEEE Conf. Comput. Vis. Patt. Recognit.*, 2016.

[50] A. Newell, K. Yang, and J. Deng, “Stacked hourglass networks for human pose estimation,” in *Proc. European Conf. Comput. Vis.*, 2016, pp. 483–499.

[51] J. Kim, J. Kwon Lee, and K. Mu Lee, “Accurate image super-resolution using very deep convolutional networks,” in *Proc. IEEE Conf. Comput. Vis. Patt. Recognit.*, 2016.

[52] B. Lim, S. Son, H. Kim, S. Nah, and K. M. Lee, “Enhanced deep residual networks for single image super-resolution,” in *Proc. IEEE Conf. Comput. Vis. Patt. Recognit. Workshops*, 2017.

[53] S. Iizuka, E. Simo-Serra, and H. Ishikawa, “Globally and locally consistent image completion,” *ACM Trans. Graphics*, vol. 36, no. 4, pp. 107–121, 2017.

[54] H. Zhao, J. Shi, X. Qi, X. Wang, and J. Jia, “Pyramid scene parsing network,” in *Proc. IEEE Conf. Comput. Vis. Patt. Recognit.*, 2017, pp. 2881–2890.

[55] P. Isola, J.-Y. Zhu, T. Zhou, and A. A. Efros, “Image-to-image translation with conditional adversarial networks,” in *Proc. IEEE Conf. Comput. Vis. Patt. Recognit.*, 2017.

[56] L. Chen, G. Papandreou, I. Kokkinos, K. Murphy, and A. L. Yuille, “Deeplab: Semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected crfs,” *IEEE Trans. Patt. Analysis Machine Intelligence*, vol. 40, no. 4, pp. 834–848, April 2018.

[57] J. Yu, Z. Lin, J. Yang, X. Shen, X. Lu, and T. S. Huang, “Generative Image Inpainting with Contextual Attention,” in *Proc. IEEE/CVF Conf. Comput. Vis. Patt. Recognit.*, 2018, pp. 5505–5514.

[58] S. Ioffe and C. Szegedy, "Batch normalization: Accelerating deep network training by reducing internal covariate shift," in *Proc. Intl. Conf. Mach. Learn.*, 2015, pp. 448–456.

[59] F. Yu and V. Koltun, "Multi-scale context aggregation by dilated convolutions," in *Proc. Intl. Conf. Learn. Represent.*, 2016.

[60] M. D. Zeiler, D. Krishnan, G. W. Taylor, and R. Fergus, "Deconvolutional networks," in *Proc. 2010 IEEE Computer Society Conf. Comput. Vis. Patt. Recognit.*, 2010, pp. 2528–2535.

[61] A. Dosovitskiy, P. Fischer, E. Ilg, P. HÃďusser, C. Hazirbas, V. Golkov, P. v. d. Smagt, D. Cremers, and T. Brox, "Flownet: Learning optical flow with convolutional networks," in *Proc. 2015 IEEE Intl. Conf. on Computer Vision*, 2015, pp. 2758–2766.

[62] S. Meister, J. Hur, and S. Roth, "UnFlow: Unsupervised learning of optical flow with a bidirectional census loss," in *Proc. AAAI Conf. Artificial Intell.*, 2018.

[63] D. Sun, X. Yang, M.-Y. Liu, and J. Kautz, "PWC-net: CNNs for optical flow using pyramid, warping, and cost volume," in *Proc. 2018 IEEE/CVF Conf. Computer Vision and Pattern Recognition*, July 2018, pp. 8934–8943.

[64] E. Simoncelli and W. Freeman, "The steerable pyramid: a flexible architecture for multi-scale derivative computation," in *Proc. Intel. Conf. on Image Processing*, vol. 3, 1995, pp. 444–447.

[65] D. G. Lowe, "Distinctive image features from scale-invariant keypoints," *Intl. Journal of Computer Vision*, vol. 60, no. 2, pp. 91–110, 2004.

[66] Z. Wang, A. Bovik, H. Sheikh, and E. Simoncelli, "Image quality assessment: from error visibility to structural similarity," *IEEE Trans. Img. Process.*, vol. 13, no. 4, pp. 600–612, 2004.

[67] O. Ronneberger, P.Fischer, and T. Brox, "U-net: Convolutional networks for biomedical image segmentation," in *Proc. Medical Image Computing and Computer-Assisted Intervention*, 2015, pp. 234–241.

[68] M. Bertalmio, G. Sapiro, V. Caselles, and C. Ballester, "Image inpainting," in *Proc. the 27th Annual Conf. Comput. Graphics and Interactive Techniques*, 2000, pp. 417–424.

[69] A. Criminisi, P. Perez, and K. Toyama, "Region filling and object removal by exemplar-based image inpainting," *IEEE Trans. Image Process.*, vol. 13, no. 9, pp. 1200–1212, 2004.

[70] D. Pathak, P. Krähenbühl, J. Donahue, T. Darrell, and A. A. Efros, "Context Encoders: Feature Learning by Inpainting," in *Proc. IEEE Conf. Comput. Vis. Patt. Recognit.*, 2016, pp. 2536–2544.

[71] D. Yu, H. Wang, P. Chen, and Z. Wei, "Mixed pooling for convolutional neural networks," in *Proc. International conference on rough sets and knowledge technology*, 2014, pp. 364–375.

[72] B. Zhou, A. Lapedriza, A. Khosla, A. Oliva, and A. Torralba, "Places: A 10 million

image database for scene recognition," *IEEE Trans. Patt. Analysis and Machine Intell.*, vol. 40, no. 6, pp. 1452–1464, 2018.

[73] K. Karaduzovic-Hadziabdic, J. H. Telalovic, and R. Mantiuk, "Expert evaluation of deghosting algorithms for multi-exposure high dynamic range imaging," in *Proc. HDRi2014-Second Intl. Conf. and SME Workshop HDR Imag.*, 2014.

[74] O. T. Tursun, A. O. Akyüz, A. Erdem, and E. Erdem, "An objective deghosting quality metric for hdr images," in *Proc. the 37th Annual Conf. of the European Association for Comput. Graphics*, 2016, pp. 139–152.

[75] HDRsoft Ltd., "Photo editing software for hdr and real estate photography | photomatix," https://www.hdrsoft.com/ (2021/1/18 参照).

[76] M. D. Grossberg and S. K. Nayar, "Modeling the space of camera response functions," *IEEE Trans. Patt. Analysis and Machine Intell.*, vol. 26, no. 10, pp. 1272–1282, 2004.

関連論文の印刷公表の方法および時期

4 章および 5 章に関連

(1) Isana funahashi，Taichi Yoshida，Zhang Xi and Masahiro Iwahashi，“Image Adjustment for Multi-exposure Images Based on Convolutional Neural Networks,” IEICE Transactions on Information and Systems，Vol.E105-D，No.1，pp.123–133，Jan. 2022.（2021 年 9 月採録決定済）