Vol.2020-CG-180 No.30
Vol.2020-DCC-26 No.30
Vol.2020-CVIM-223 No.30
2020/11/6

非拡散反射成分を考慮した人物全身画像の再照明

田島 大地[†1,a)]　金森 由博[†1,b)]　遠藤 結城[†1,c)]　三谷 純[†1,d)]

概要：画像中の人物の再照明のために，被写体の形状，反射率および照明をニューラルネットワークによって推定する手法が提案されている．しかし，学習用データとして利用可能な，一般に市販されている 3D 人物モデルの多くは，テクスチャとして拡散反射成分しか持たず，非拡散反射成分の学習が困難であった．そこで本研究では，非拡散反射も含めた再照明の手法を提案する．拡散反射限定の既存手法を実写人物画像に適用し，出力から入力画像を再構成したのち，再構成画像と入力画像との差分を求める．この差分を再現するように新たなニューラルネットワークを導入し訓練する．これにより，拡散反射限定の既存手法よりも写実的な再照明を実現する．

Relighting of Full-Body Human Images with Non-Diffuse Reflective Components

Abstract: An existing technique enables the relighting of full-body human images via estimation of the shape, reflectance, and illumination using neural networks. However, it cannot handle non-diffuse reflective components because its training dataset is synthesized from mostly diffuse-only commercial 3D human models. In this paper, we propose a method for integrating non-diffuse reflective components into the relighting pipeline. Our key idea is to add another neural network for learning the differences between photographs and diffuse-only reconstruction by the existing technique. By adding the difference on top of the diffuse-only reconstruction, we demonstrate that our method accomplishes more realistic relighting with non-diffuse reflection.

1. はじめに

照明は，ポートレート写真における被写体の印象に大きな影響を与える．撮影スタジオのような環境であれば，照明を操作することで被写体の印象を自在に変えることができるが，屋外など照明の制御が難しい環境下では撮影の自由度が制限される．もし被写体の陰影を後から修正できれば，ポートレート写真の撮影後であっても自在に理想の見た目に仕上げることができる．このように，一度撮影された被写体が別の照明環境下でどう見えるかを再現する技術を再照明と呼ぶ．

着衣の人物全身画像を対象とした再照明の手法 [1] が提案されている．この手法は，畳み込みニューラルネットワーク（convolutional neural network; CNN）を用い，反射率，形状および照明を推定したのち，推定された照明を新たな照明に差し替えることにより再照明を実現する．しかしこの手法では，髪や肌に存在する光沢成分などの非拡散反射成分が十分に再現できず，ツヤのないマットな質感の画像が生成されてしまう．その原因として，学習に用いられる市販の 3D 人物モデルの多くがテクスチャとして光沢成分を含まず，拡散反射成分しか持たないので，拡散反射以外の成分を学習させるのが難しいことが挙げられる．一方，LightStage と呼ばれる大型の装置を用いて撮影された，非拡散反射成分を含む学習データを用いる手法も存在するが，そのようなデータは一般に公開されていない．

そこで本研究では，人物全身画像を入力とした，より写実的な再照明画像を生成する新たな手法を提案する．非拡散反射成分を考慮するために，提案手法では (1) 拡散反射のみを考慮した既存手法 [1] を適用して再構成画像を得た後，(2) 再構成画像と実写画像との差分を再現できるように別途ネットワークを学習させる．既存手法の出力と，新たなネットワークが出力する差分を足し合わせることによ

[†1] 筑波大学
University of Tsukuba, Tennoudai 1-1-1, Tsukuba, Ibaraki, 305–8573, Japan
[a)] tajima@cgg.tsukuba.ac.jp
[b)] kanamori@cs.tsukuba.ac.jp
[c)] endo@cs.tsukuba.ac.jp
[d)] mitani@cs.tsukuba.ac.jp

Vol.2020-CG-180 No.30
Vol.2020-DCC-26 No.30
Vol.2020-CVIM-223 No.30
2020/11/6

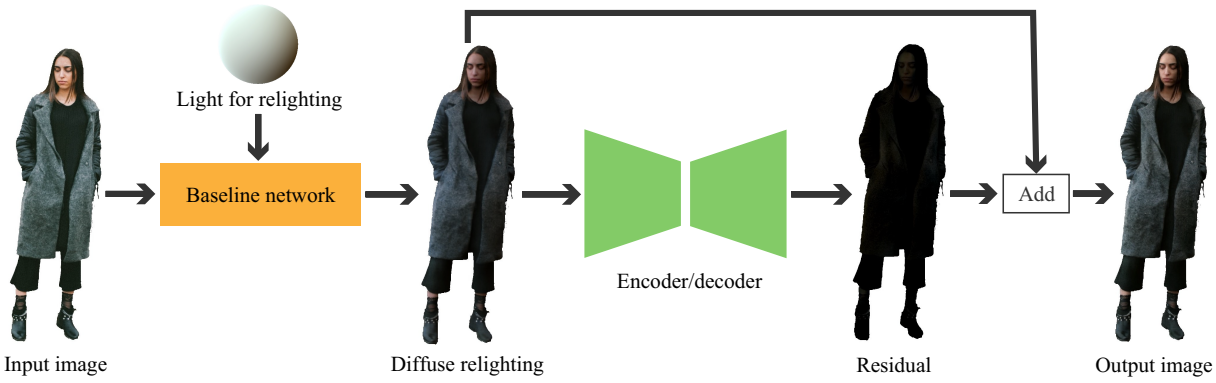


図 1 提案手法を用いた再照明の流れ．ベースライン手法によって得られた拡散反射のみの再照明結果に，差分となる非拡散反射成分を新たな CNN で推定して加え，再照明結果を出力する．

り，拡散反射のみを考慮した手法よりも，より現実に近い結果が得られることを示す．

2. 関連研究

単一画像を入力として物理ベースで再照明を行うものとして，人物の顔画像を対象とした研究が多数行われている．例えば Sengupta らの研究 [2] では，CNN によって顔の形状と反射率，照明を推定し，推定した照明を新たな照明に差し替えることによって再照明を行っている．この再照明手法は人物全身画像にも適用することができる．しかし，光の遮蔽は考慮されていないので遮蔽が起こる部位が不自然に明るくなってしまう．また，拡散反射のみを考慮しているため，それ以外の反射成分を再現することは難しい．

Zhou らの研究 [3] では陽に照明計算を行わず，CNN によって再照明計算を近似している．再照明時には肌にある光沢を再現できている．この研究ではデータセット作成時に，変形可能な 3D 頭部モデルを実写顔画像にフィッティングさせることで顔の形状データ (法線マップ) を得ている．しかし今回対象とする着衣の人物全身画像では，実写画像に対して複雑な衣服の形状まで含めて 3D モデルをフィッティングすることは難しい．

LightStage を使って生成した正解データを用いることで写実的な再照明を行う手法も提案されている [4] [5]．LghtStage とは，全方位に複数の光源とカメラを配置した球形ドーム型キャプチャーシステムであり，再照明のための非拡散反射成分も含む高品質な学習データを得ることができる．しかし，そのようなデータセットは一般に公開されておらず，本研究で学習に用いることはできない．

3. ベースライン手法

人物全身画像の再照明手法である Kanamori と Endo の研究 [1] を本研究のベースライン手法として採用した．他手法との違いとして，Zhou らの研究 [3] では陰影計算時に形状による光の遮蔽が陽に考慮されていないが，この手法 [1] では光の遮蔽情報を含む光伝達マップを直接推定することで，服の皺，脇や股など凹んだ部分での光の遮蔽を陽に扱える．この手法では，1 つのエンコーダと 3 つのデコーダからなるマルチタスク型ネットワークを用いて，再照明に必要な情報である，反射率と，光の遮蔽情報を含む光伝達マップ，そして照明情報を推定する．学習データセットは市販の 3D 人物モデルをレンダリングして作成しているが，テクスチャとして光沢などの非拡散反射成分を含まないため，出力は拡散反射に限定されている．

4. 提案手法

提案手法の入力はベースライン手法 [1] と同様，マスク処理された人物全身画像 (これに加えて再照明時には球面調和関数で表現された光源情報) であり，出力として非拡散反射成分を考慮した再照明画像を得る．提案手法では，ベースライン手法 [1] で再照明画像を得た後，その被写体の非拡散反射成分を推定し再照明するという 2 段階の方法で行う．具体的には，まず入力画像にベースライン手法を適用して，拡散反射成分のみを考慮した再照明画像を得る．このとき，拡散反射のみではどうしても実写の入力画像との差分が生じるため，次に再照明画像と実写画像との差分を出力する別の CNN に入力する．こうして得られた差分を，ベースライン手法で得られた再照明画像と足し合わせることで，非拡散反射成分を考慮した再照明画像を得る．**図 1** に本手法の流れを示す．

訓練，テスト時には人物全身画像とその二値マスクを必要とする．本研究で用いる学習データセットについては 4.3 節で説明する．

4.1 ネットワークモデル

本研究ではベースライン手法で使われているネットワークを改変したものを使う．残差推定のみを行うため，デコーダは 3 つではなく 1 つのみ使用し，エンコーダとデコーダをスキップ接続でつなぐ．新たな設計として，以下

Vol.2020-CG-180 No.30
Vol.2020-DCC-26 No.30
Vol.2020-CVIM-223 No.30
2020/11/6

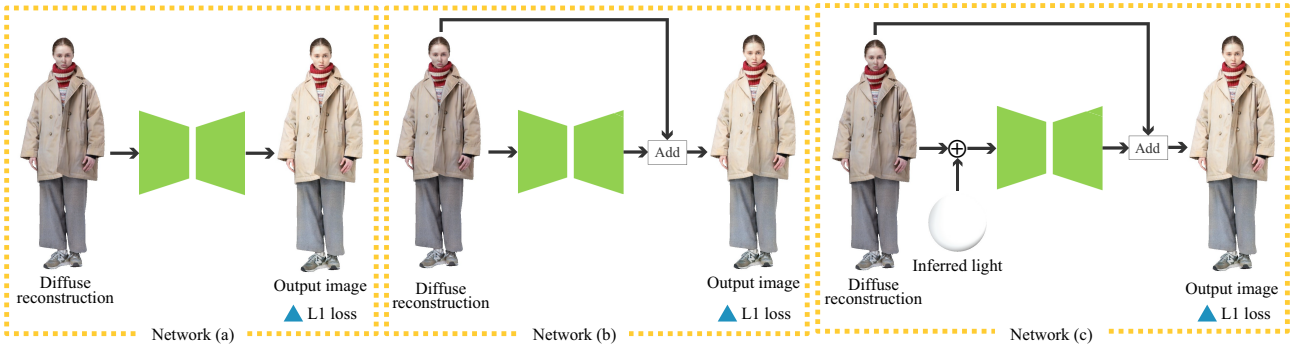


図 2 試みた 3 つのネットワーク構造.

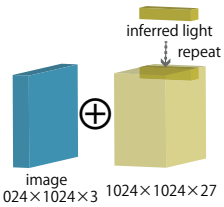
の 3 つのネットワーク構造を試みた.

(a) ネットワークの出力がそのまま非拡散反射成分を考慮した人物画像となるようなネットワーク (**図 2** の左)

(b) ネットワークの出力が再構成画像と正解画像の差分になるようなネットワーク (図 2 の中央)

(c) ネットワークの入力として，推定された光源も用いるネットワーク (図 2 の右)

上記のうち (a) は，ネットワークの出力を正解画像に近づけるように学習させる最も単純なネットワーク構造である．ネットワークの出力がそのまま非拡散反射成分を考慮した出力になるようにした．(b) は，ResNet [6] のように，入力である拡散反射のみ考慮した再構成画像をネットワークの出力と足し合わせることで，ネットワークの出力が入力画像と正解画像の差分になるように設計した．(c) は，(b) の構造に加えて，入力として推定された光源も用いるネットワーク構造である．光源は，右上挿入図に示すように入力画像と同じ空間分解能を持つように空間的に繰り返すことで，入力画像と連結させたものを入力とする．これは，光源の指向性をよく学習することを狙った設計である．我々の予備実験の結果，定量的・定性的に (b) が最も優れていたため，以下の実験では (b) を採用する．



4.2 ネットワークの学習

提案手法では，まずベースライン手法のネットワークを事前に学習させておく．そしてそのネットワークパラメータは固定した状態で実写画像から拡散反射成分のみの再構成画像を得て，新たなネットワークで再構成画像と実写画像との差分を訓練させる．損失関数には，最終的な出力について正解との L1 損失を用いた．

また光源に関して，ベースライン手法では主に屋内で計測された 40 個の光源を学習に用いていたが，本研究では屋内と屋外の両方を含む 3,450 個に拡張した．ベースライン手法では学習時に人物データを 1 つ選ぶたびに全ての光源データを使って学習させていたが，この方法では光源データのループの間は同じ人物データを使うことになり，学習が偏る可能性がある．本研究では，人物データ 1 つに対してランダムに光源を 1 個選択して学習するように変更した．以降，光源データを拡張して学習させたモデルをベースライン手法とする．

図 3 本研究で用いる学習データの例．各人物データについて，二値マスクとのペアからなる．

4.3 データセットの作成

本手法で使う人物全身画像データは，立ちポーズに限定し，ファッション系 Web サイトから取得した．データセットには屋内と屋外で撮影された実写の人物全身画像と，それに対応する二値マスクが 9,152 組含まれている．これらのうち 8,900 体分を訓練データ，252 体分をテストデータとした．マスクの作成には商用の画像切り抜きサービス [7] を用いた．画像および二値マスクの解像度は 1024 × 1024 画素とした．**図 3** に学習データの例を示す．

5. 実験結果

提案手法を Python および PyTorch を用いて実装し，NVIDIA GeForce GTX 1080 Ti 上で学習・推論を行った．最適化には Adam を用い，モーメント推定に使う指数減衰率は {0.5,0.999} とした．学習率は CosineAnnealing を用い，1 周期を 40 エポックとしてスケジューリングを行った．バッチサイズは 1 とした．訓練にかかった時間は 1 つの GPU を用いて 1024 × 1024 画素のデータを入力した場合，1 エポックあたり約 30 分であった．実験で用いた本手法のモデルは，ほぼ学習が収束したとみられる 180 エポックまで学習したものである．推論にかかる時間は 1024 × 1024

Vol.2020-CG-180 No.30
Vol.2020-DCC-26 No.30
Vol.2020-CVIM-223 No.30
2020/11/6

表 1　入力画像を再構成したときの RMSE および SSIM．RMSE は二値マスク内の画素のみを使用し，SSIM はバウンディングボックス内の画素を使用した．

	RMSE	SSIM
ベースライン手法	0.0712	0.948
提案手法	**0.0561**	**0.957**

画素の入力 1 つあたり 0.18 秒程度である．

入力画像の再構成画像と再照明結果について，ベースライン手法と提案手法との比較を行った．定量評価の際には，再照明時に使う光源で照明された実写の正解データが存在しないため，入力画像に対する再構成結果の精度比較も行った．以下の実験に用いたデータはテストデータであり，訓練データは用いていない．

5.1　入力画像を再構成したときの比較

5.1.1　定量的比較

ベースライン手法と提案手法を使って，本研究で用意したテスト用データセット 252 枚の再構成結果の RMSE と SSIM の値を**表 1** に示す．RMSE は各画像の二値マスク外の影響を除外するため，二値マスク内の画素のみを使って計算した．定量評価では，いずれもベースライン手法より提案手法の結果の方がよく，入力である実写画像に近づいていることがわかる．

5.1.2　定性的比較

ベースライン手法と提案手法の推定結果の，定性比較の結果を**図 4** に示す．図 4 の右端から 2，3 番目は正解データである入力画像との輝度差 (絶対値誤差) を可視化したものである．色が白色に近いほど入力 (正解) 画像との誤差が少なく，黒に近いほど誤差が大きいことを示す．提案手法の方が入力画像との差が少なくなっており，より精度の高い再現ができていることがわかる．また，非拡散反射成分が多く存在すると考えられる肌などの部位がうまく学習できているか確認するため，ベースライン手法と提案手法の差分も可視化した (図 4 の右端)．提案手法による出力の画素値からベースライン手法の出力の画素値を引いた時の正負を，それぞれ赤色，青色の濃淡で示す．ベースライン手法にはない額や腕などの部位のハイライトが復元できていることがわかる．

5.2　再照明時の比較

再照明時の比較では，異なる照明で再照明された実写画像の正解データが存在しないため，定量的に評価を行うことができず，視覚的な結果についてのみ比較する．再照明時の比較を**図 5** に示す．ベースライン手法による再照明結果は全体的にマットな質感であるのに対して，提案手法ではより現実に近い陰影が付加されている．下段は，提案手法からベースライン手法で得られた結果を引いた差分を表している．入力画像から光沢がつくと予想される額部分のハイライトが移動しているのがわかる．このことから，肌にある非拡散反射成分を考慮できていると言える．

6.　まとめと今後の課題

本研究では，着衣の人物全身画像を対象とする，非拡散反射成分を考慮した再照明手法を提案した．ベースライン手法で得られた再照明画像をより写実的に再現するために，実写と拡散反射のみを考慮した画像の差分を学習させるようなネットワークを提案した．その結果，肌などににある非拡散反射成分を考慮した写実的な出力を得ることができた．再照明画像については実写正解データが存在しないため，定量的な改善は確認できなかったが，ベースライン手法よりも視覚的に良好な結果が得られることを確認できた．今後の課題として，人物が映った動画に適用できるよう提案手法を拡張したい．

謝辞　本研究は JSPS 科研費 JP19H04130 の助成を受けたものです．

参考文献

[1] Yoshihiro Kanamori and Yuki Endo. Relighting humans: occlusion-aware inverse rendering for full-body human images. *ACM Trans. Graphics*, Vol. 37, No. 6, pp. 1–11, Dec 2018.

[2] Soumyadip Sengupta, Angjoo Kanazawa, Carlos D. Castillo, and David W. Jacobs. SfSNet: Learning shape, reflectance and illuminance of faces 'in the wild'. In *CVPR 2018*, June 2018.

[3] Hao Zhou, Sunil Hadap, Kalyan Sunkavalli, and David W. Jacobs. Deep single-image portrait relighting. In *ICCV 2019*, October 2019.

[4] Tiancheng Sun, Jonathan T. Barron, Yun-Ta Tsai, Zexiang Xu, Xueming Yu, Graham Fyffe, Christoph Rhemann, Jay Busch, Paul Debevec, and Ravi Ramamoorthi. Single image portrait relighting. *ACM Trans. Graphics*, Vol. 38, No. 4, pp. 1–12, Jul 2019.

[5] Thomas Nestmeyer, Jean-Francois Lalonde, Iain Matthews, and Andreas Lehrmann. Learning physics-guided face relighting under directional light. In *CVPR 2020*, June 2020.

[6] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. In *CVPR 2016*, June 2016.

[7] remove.bg. `https://www.remove.bg` [accessed 12 October 2020].

Vol.2020-CG-180 No.30
Vol.2020-DCC-26 No.30
Vol.2020-CVIM-223 No.30
2020/11/6

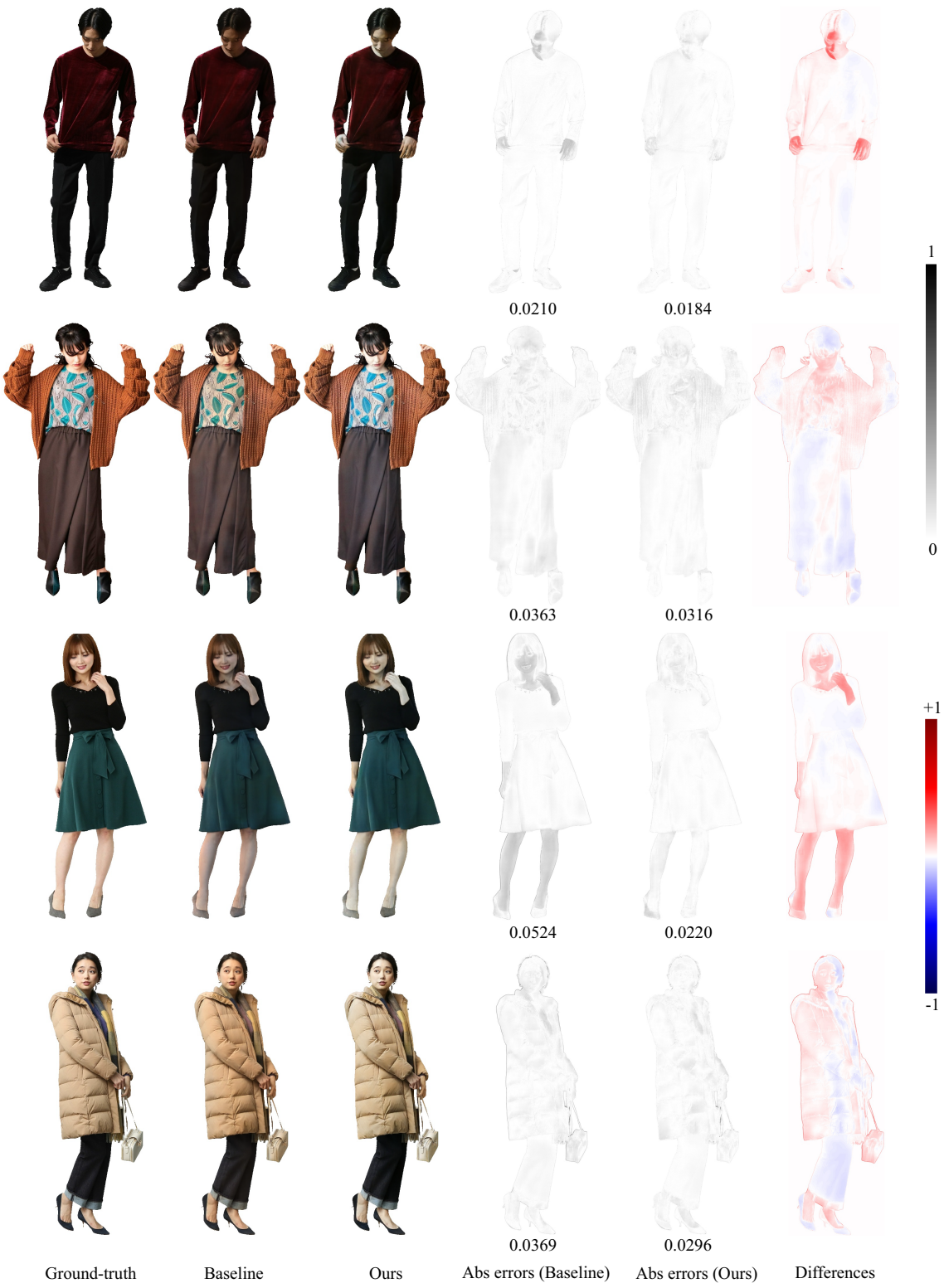


図 4　入力画像を再構成した際の定性的比較.

Vol.2020-CG-180 No.30
Vol.2020-DCC-26 No.30
Vol.2020-CVIM-223 No.30
2020/11/6

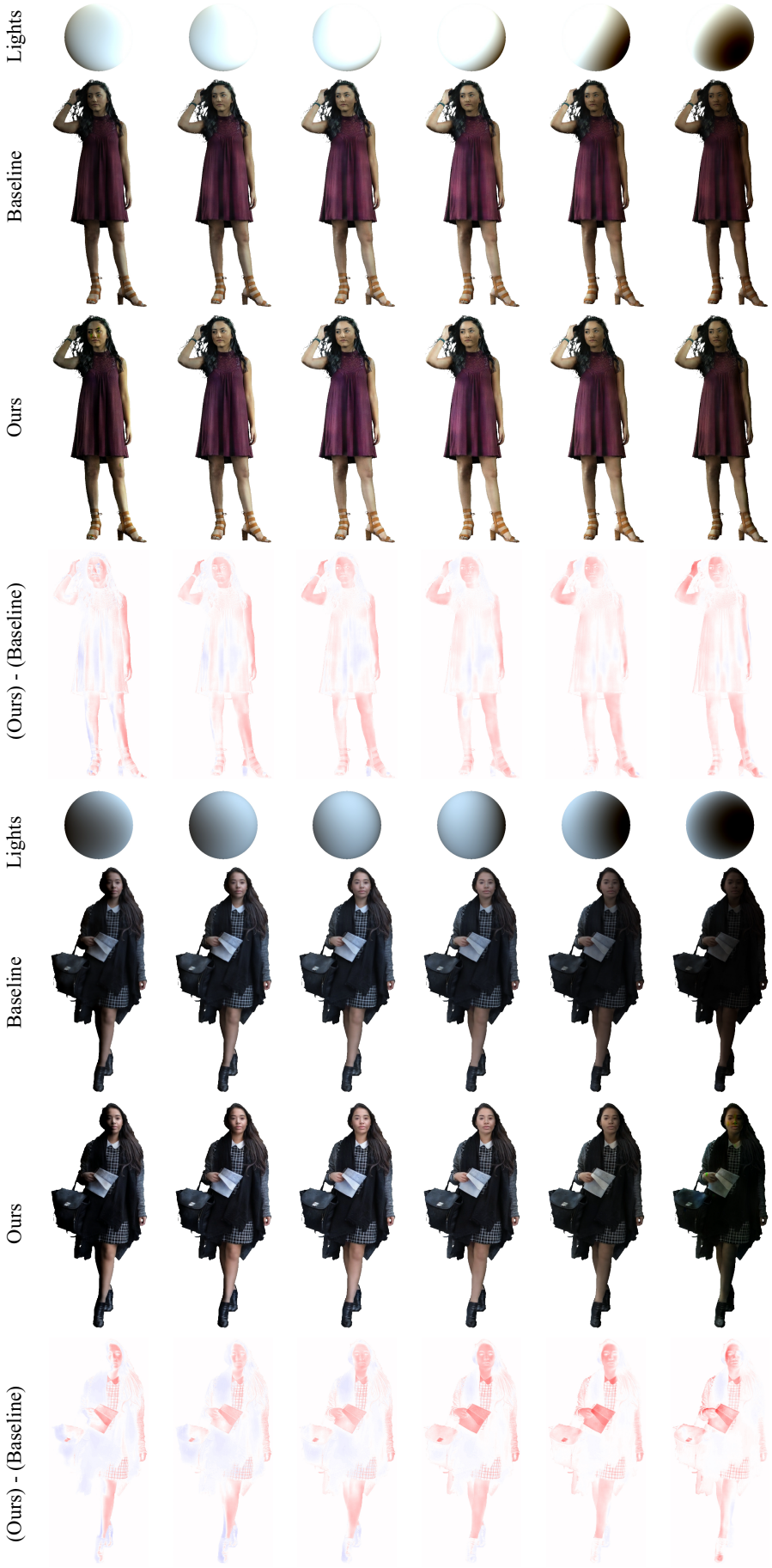


図 5　再照明時の定性的比較.