Vol.2022-CVIM-230 No.18
2022/5/12

球面畳み込みニューラルネットワークを用いた光源状況推定

延與 唯人[1]　山下 浩平[1]　延原 章平[1]　西野 恒[1]

概要：物体の見えからの光源状況推定は，状況認識や自然な拡張現実に不可欠である．この問題は不良設定な逆問題であり，ニューラルネットワークによって学習できる解の拘束が有効である．しかし，平面畳み込みを用いる既存手法では，球面上の信号である光源状況を平面上のものとして扱っている．ドメインを無視したこのような処理では，フィルターの歪みや信号の連続性の破綻が生じるため，球面の信号や関数を正しく表現できない．本研究では，ドメインを正確に扱うために，球面畳み込みを用いたニューラルネットワークでこの問題を解く．既存手法と比較し，提案手法がより少ないパラメーターで，精度高く推定できることを示した．

1. 光源状況推定

ある環境の光源状況を推定することは，状況認識や拡張現実などを行う上で欠かせない技術である．光源状況がわかれば，置かれた場所や状況を認識したり，その場にある物体の反射特性を推定したりできる．また，仮想物体がその環境の光源でレンダリングされた，自然な仮想現実を提供することもできる．観測している物体を中心とした光源状況を直接的に観測するには，その同じ場所に全天球カメラや鏡面球を設置する必要がある．一方で物体の見えや形状は RGBD カメラなどで俯瞰的に観測できる．そこで本研究では，形状既知の物体の画像から，対象の物体を中心とした光源状況を推定することを目的とする．この問題は，物体画像が形状・反射特性・光源状況という複数の要素からなるものであるため不良設定な逆問題である．

物体画像から光源状況を推定する既存手法には，物理的な拘束を考慮する手法や膨大な訓練データによってその拘束を学習する手法がある．物理的な拘束を明示的に与える方法では，光源状況を単純なものと仮定するか事前分布を定式化する必要があり，得てして表現力の低いものとなってしまう．演繹的にモデルを立てることは現実的ではないため，ニューラルネットワークによって大量のデータで帰納的に拘束を学習する手法が有効である．しかし，既存の研究では二次元平面の畳み込みニューラルネットワークを用いており，光源状況などが方向に対する関数という事実を考慮していない．本来は球面上の信号を平面に投影して扱うと，連続している部分が途切れたり，フィルターの形状が実際の球面上で歪んだりする．これは球面信号を出力する点においても，球面関数の演算を表現する点においても，不正確な表現でしかない．

そこで本研究では，球面畳み込みにより構成されたニューラルネットワーク（球面 ConvNet）を用いて光源状況に対する拘束を学習により体得する．ネットワークは全体として，物体画像と形状から近似的に得られる，表面の法線方向と見えの対応・反射マップを入力に，光源状況として鏡面球の反射マップを出力するものである．前半では直接得られるスパースな反射マップを敵対的損失も導入して密な反射マップに変換する．後半では学習によって体得した光源状況に対する拘束によって，反射特性の逆関数のようにネットワークを作用させ密な反射マップから鏡面球の反射マップを出力する．

物体の合成画像や実画像からシーンの光源状況を推定する実験を行い，提案モデルの評価を定性的・定量的に行った．訓練されたネットワークを用いて，異なるデータセットに対して手法を適用し評価した．訓練には，光源状況と反射特性を組み合わせて作製した合成データセットを用いた．その結果，球面畳み込みがより少ないパラメーターで，かつ推定時には最適化を要せずに，より自然な精度の高い推定ができることがわかった．また，本研究は球面 ConvNet を生成モデルに用いた初めての研究であり，球面畳み込みが生成モデルにも応用できることを示した．我々は球面畳み込みを用いた光源状況推定を通して，光源状況の球面という本質を捉え，球面畳み込みの可能性を広げた．

[1] 京都大学大学院情報学研究科

© 2022 Information Processing Society of Japan

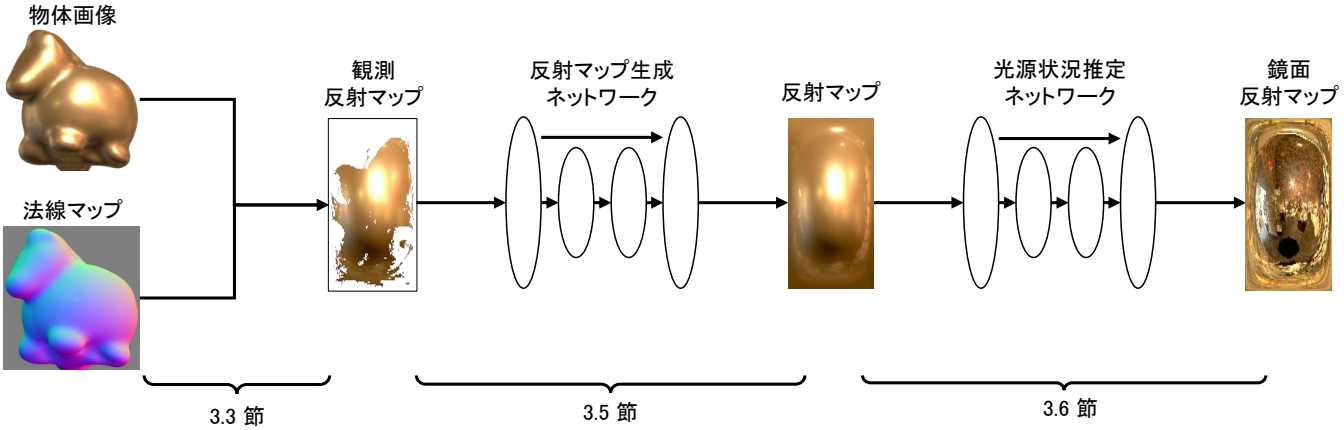

図 1 提案手法の概要図．物体画像と法線マップを反射マップ空間に変換する．得られる欠損のある観測反射マップから完全な反射マップを復元する．復元結果を元に光源状況として鏡面反射マップを推定する．

2. 関連研究

物体画像から光源状況を推定する手法として，物理的な拘束を考慮するものには，Lombardi と Nishino らによるベイズ推定に基づいた手法がある [1][2]．彼らは最大事後確率推定（MAP 推定）を導入して，物体の観測画像が与えられたときの事後確率が最大となるような光源状況と物体の反射特性を求めた．MAP 推定を行うには，光源状況と反射特性をモデル化し事前分布を定義する必要がある．彼らは光源状況の事前分布に，エントロピーと勾配を用いた確率分布関数を定義した．しかし，このような人工的に定式化された事前分布は，結果的に，なめらかすぎる光源状況の推定につながってしまうという問題がある．

近年では，深層畳み込みニューラルネットワークを用いた光源状況推定手法が高い精度を出している．Georgoulis らは畳み込みニューラルネットワークを用いて，物体の画像から直接的に光源状況を推定する手法を提案した [3]．彼らは大量のデータで事前学習を行うことで，各ネットワークに形状や反射特性，光源状況の事前分布を埋め込み，推定を可能にした．Legendre らは，屋内外の様子を切り取った画像や動画から，周囲の光源状況を推定する手法を提案した [4]．彼らはいくつかの素材の球をカメラの前に固定し，様々な環境で背景と同時に撮影したデータセットを用意した．推定光源から生成した球の画像と実際の球の画像の差を損失として畳み込みニューラルネットワークを学習させた．

畳み込みニューラルネットワークを，従来の手法と融合させた研究もある．Chen らは，Lombardi らの MAP 推定手法 [2] を踏襲しつつ，Deep Image Prior[5] を用いることで，人工的に事前分布を定式化せずに推定を行う手法を提案した [6]．光源状況を畳み込みニューラルネットワークの出力とすることで，Deep Image Prior の自然画像に対する特性を光源状況の事前分布として用いた．光源状況自体にではなくネットワークの内部パラメーターに対して MAP 推定を行うことで，光源状況の事前分布を人工的に定式化することなく構造的に埋め込むことを達成した．

しかし以上の畳み込みを用いた既存の手法は，光源状況のような球面上の信号を平面に投影し，単純な平面画像として扱っている．これは平面から実際の球面上に逆投影したとき，フィルターが歪んでしまうことを意味し，信号や関数の表現として問題がある．そこで本研究では球面 ConvNet を用いて光源状況を推定する．

3. 球面畳み込みを用いた光源状況推定

本節では，まず提案手法の概要を俯瞰した後，物体の見え方の定式化を行う．その後，物体の画像と形状から光源状況を推定する提案手法を三つの段階に分けて詳しく述べる．

3.1 推定手法の概要

図 1 に，提案する光源状況推定手法の概要を示す．本手法で入力とするのは，物体の画像と，画像の各ピクセルに写る物体表面の法線方向を表す法線マップである．画像と法線マップを元に，法線方向に対する輝度値という関係を表す反射マップに変換する．反射マップの定義域は本来半球となるが，物体の法線はそのすべての方向を持つとは限らないため，足りない情報を補って完全な反射マップを生成する．復元された反射マップを元にして，最終的に光源状況を推定する．入力に対し，この三つの段階の処理を行った出力として光源状況を推定する．各段階について，それぞれ本節の 3.3, 3.5, 3.6 節で詳しく述べる．

生成と推定の処理を行うために，本手法では球面 ConvNet を用いる．扱っている反射マップや光源状況は，方向に対して値を表すものである．球面畳み込みは，これらの信号の適切な処理になっている．この球面 ConvNet に

© 2022 Information Processing Society of Japan

ついては 3.4 節で詳しく述べる．

3.2 見えの生成モデル

観測される物体表面の輝度値は，物体自身が光源として発する光と，周囲から入射して反射・透過した光の和とみなせ，レンダリング方程式で記述できる [7]．時間変化を無視すればレンダリング方程式は，

$$L_{\mathrm{o}}(\boldsymbol{x},\boldsymbol{\omega}_{\mathrm{o}},\lambda) = L_{\mathrm{e}}(\boldsymbol{x},\boldsymbol{\omega}_{\mathrm{o}},\lambda) + \int_{\mathbb{S}^2} f_{\mathrm{s}}(\boldsymbol{x},\boldsymbol{\omega}_{\mathrm{i}},\boldsymbol{\omega}_{\mathrm{o}},\lambda) L_{\mathrm{i}}(\boldsymbol{x},\boldsymbol{\omega}_{\mathrm{i}},\lambda) |\boldsymbol{\omega}_{\mathrm{i}} \cdot \boldsymbol{n}| \mathrm{d}\boldsymbol{\omega}_{\mathrm{i}}\,, \tag{1}$$

と表される．ここで λ は波長，$\boldsymbol{x}$ は物体表面の位置，$\boldsymbol{n}$ は $\boldsymbol{x}$ での物体表面の法線方向，$\boldsymbol{\omega}_{\mathrm{o}}, \boldsymbol{\omega}_{\mathrm{i}}$ はそれぞれ法線方向 $\boldsymbol{n}$ を基準にした光の入射方向・出射方向である．$L_{\mathrm{i}}, L_{\mathrm{o}}$ はそれぞれ入射・放射する放射輝度，L_{e} は物体自身が発する放射輝度を意味する．また，f_{s} は $\boldsymbol{x}$ に $\boldsymbol{\omega}_{\mathrm{i}}$ 方向から入射した波長 λ の放射照度のうち $\boldsymbol{\omega}_{\mathrm{o}}$ 方向に出ていく放射輝度の割合を表す関数（BSDF）である．

本研究では，光源や観測者の位置，物体の材質に仮定をおく．まず周囲の光源や観測者は無限遠にあるとする．光源からの直接光だけでなく間接光も等価に扱い，方向のみに依存した放射輝度として考える．物体の大きさに対して光源や観測者がある程度遠ければ，近似的にこの仮定が成り立つ．物体の材質については，物体全体で一様であり，発光や透過がなく，入射した光が物体表面のみで反射して放射されるような素材を仮定する．またモデルを簡単にするため，物体内部で生じる相互反射や照明の遮蔽を無視することとする．

このような仮定の下では，**図 2** で示すように，レンダリング方程式を反射マップとして簡略に表現でき，物体表面の輝度値は反射マップと法線マップのみから決定できる．すなわち (1) 式のレンダリング方程式は，ある一定の方向の観測者から見ると，

$$L_{\mathrm{o}}(\boldsymbol{n},\lambda) = \int_{\Omega} f_{\mathrm{r}}\left(\boldsymbol{\omega}_{\mathrm{i}},\boldsymbol{\omega}_{\mathrm{o}}\left(\boldsymbol{n}\right),\lambda\right) L_{\mathrm{i}}(\boldsymbol{\omega}_{\mathrm{i}},\lambda) |\boldsymbol{\omega}_{\mathrm{i}} \cdot \boldsymbol{n}| \mathrm{d}\boldsymbol{\omega}_{\mathrm{i}}\,, \tag{2}$$

と表され，物体表面の法線方向 n と波長 λ のみを変数とした関数（反射マップ）となる．なお，Ω は $\boldsymbol{n}$ を中心とした半球の範囲であり，f_{s} は反射のみを扱うものとして f_{r} と新たにおいている．反射マップは物体の素材で決まる反射特性 f_{r} と周囲の光源状況 L_{i} で決まる．この反射マップを法線マップに対応付けていくことで，おおよそ物体の見えが決まる．観測者側から物体の反対側は見えないため，法線方向 $\boldsymbol{n}$ の定義域は観測者の方向の半球となる．

本研究の目的は，観測された輝度値と法線マップが既知，つまり L_{o} が部分的に得られているとき，反射特性 f_{r} が未知であるという設定のもと，光源状況 L_{i} を推定することである．(2) 式より，反射マップからの逆問題を解くことで光

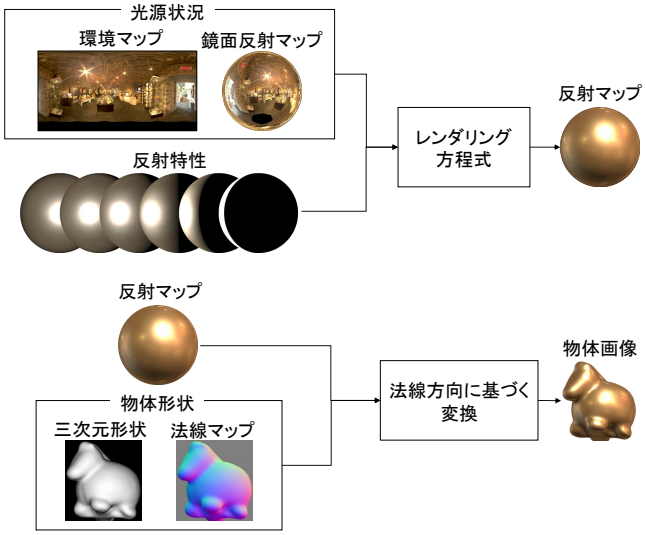

図 2 物体の見えの生成モデル．物体の見えは反射マップと本線マップで決まり，反射マップは物体の反射特性と光源状況からレンダリング方程式により決定される．

源状況を推定することができる．反射マップは (2) 式のように光源状況と反射特性の球面上の積分であり，球面畳み込みとこの点で同じであるため，本研究では球面 ConvNet を用いて，この逆関数を表現する．

3.3 観測反射マップ

ネットワークによって学習されたレンダリング方程式の逆関数を用いて光源状況を推定するには，物体画像と対応する法線マップを反射マップに変換する必要がある．**図 3** で示すように，物体形状として法線マップを用いることで，直接反射マップの $\mathbb{S}^2$ 空間に変換できる．しかし，反射マップにする上では二つのことに注意しなければならない．一つは，反射マップが実際の処理では離散化されたものであるということである．反射マップは，縦軸・横軸をそれぞれ法線方向の天頂角・方位角に対応させた二次元マップにより，等間隔に離散化して表現する．したがって，周辺の観測点を用いた補間が必要になる．もう一つは，物体画像が実際には，物体内での相互反射や照明の遮蔽を含んだものであるということである．反射マップとしては，できるだけこのような部分を使わないようにする必要がある．二つの注意点を考慮して，反射マップへの変換に以下の制約を設ける．

- 反射マップのある方向の値は，$\mathbb{S}^2$ 空間に置いて周辺にあり，かつ物体画像中において近接する画素の補間とする．
- 補間の候補に複数の選択肢がある場合は，最も明るいものを採用する．

前者の制約により，物体の角のように補間すべきではない組合せで補間してしまうことを避けつつ，遮蔽を受けた暗い部分と受けていない明るい部分を混同した補間も回避

© 2022 Information Processing Society of Japan

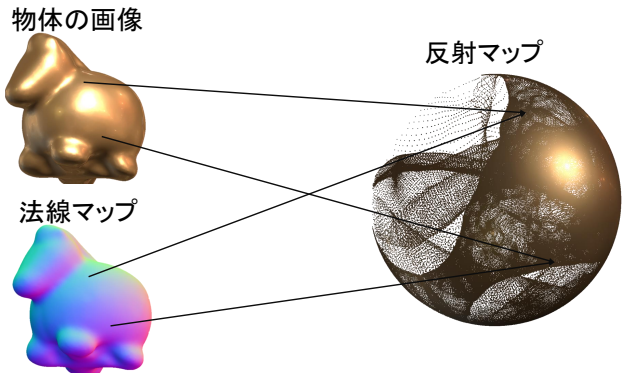

図 3 画像から反射マップへの変換．画像の各画素が $\mathbb{S}^2$ 空間で一つの点に対応する．

する．後者の制約は，相互反射や物体自身による照明の遮蔽による影響をできるだけなくすために設ける．二つのうち影響が大きい要素は物体自身によって照明が遮蔽され，物体の中に影が出てしまうことであると考えられる．そこで，明るい場所を採用することで，影となった場所をできるだけ使わないようにする．

観測画像から制約を設けて得られる反射マップは欠損を含んだスパースなものとなっている．物体の法線方向には，形状によって存在しない方向もあるからである．そのため，近傍の値で補間を施したあとでも，得られる反射マップは完全なものではない．そこで，欠損部の補間が必要になる．

3.4 球面畳み込みニューラルネットワーク

球面信号である反射マップを正しく扱うため，Cohen らの研究 [8] による球面畳み込みを用いる．球面空間 $\mathbb{S}^2$ は球面座標の天頂角 $\beta \in [0, \pi]$ と方位角 $\alpha \in [0, 2\pi]$ の二変数で表現できる．チャネル数 K の球面実信号は $f : \mathbb{S}^2 \to \mathbb{R}^K$ と表現される．例えば色を RGB で表せば $K = 3$ となる．以下では簡単のため $K = 1$ とする．畳み込むフィルターもまた同じように球面実信号で表現できる．

二次元平面の畳み込みではフィルターを縦横の並進によりずらすことから類推して，球面信号同士の畳み込みではフィルターを回転によってずらすと考えることができる．三次元空間における回転は自由度 3 の $\mathbb{SO}(3)$ 多様体であり，Z-Y-Z オイラー角 $\alpha \in [0, 2\pi], \beta \in [0, \pi], \gamma \in [0, 2\pi]$ で一意に表現できる．球面信号 f を $R \in \mathbb{SO}(3)$ で回転させる演算子を L_R とすると回転後の球面信号 $L_R f$ は，$[L_R f](x) = f(R^{-1}x)(x \in \mathbb{S}^2)$ と書ける．また，二つの $\mathbb{S}^2$ 信号 ψ, f の内積は，$\langle \psi, f \rangle = \int_{\mathbb{S}^2} \psi(x) f(x) \mathrm{d}x$ と定義できる．以上から，ψ, f の畳み込みは

$$[\psi \star f](R) = \langle L_R \psi, f \rangle = \int_{\mathbb{S}^2} \psi(R^{-1}x) f(x) \mathrm{d}x \,, \quad (3)$$

と定義される．

$\mathbb{S}^2$ 信号の畳み込み結果は $\mathbb{SO}(3)$ 信号となるので，複数層からなる球面 ConvNet を構築するため，$\mathbb{SO}(3)$ に対しても畳み込みを定義する必要がある．$\mathbb{SO}(3)$ に対するフィルターのずらし方を，$\mathbb{S}^2$ 信号のとき同様，$\mathbb{SO}(3)$ の回転 R で定義する．二つの $\mathbb{SO}(3)$ 信号 ψ, f の畳み込みは，簡単のために $K = 1$ とすると，

$$[\psi \star f]((R) = \int_{\mathbb{SO}(3)} \psi(R^{-1}Q) f(Q) \mathrm{d}Q \,, \quad (4)$$

で定義される．$x \in \mathbb{S}^2$ の回転 Rx が x 自体の回転を表現しているのに対し，$Q \in \mathbb{SO}(3)$ の回転 RQ は実際には二つの回転の合成になっている．

以上の計算を，実際の数値計算で扱える離散化された信号に対しては，周波数領域に変換して行う．球面の場合，回転した後，画素の格子が全体で揃うような回転と球面の離散化の組合せが存在しない．そのため (3), (4) 式で定義される畳み込みを，一般化されたフーリエ変換（GFT）を用いて周波数領域での積によって計算する．一次元信号に対するフーリエ変換が基底関数 $\exp(\mathrm{j}nx)$ への射影であるように，$\mathbb{S}^2$ 信号 $f : \mathbb{S}^2 \to \mathbb{R}^K$ に対する GFT は，球面調和関数 $Y_m^l(x)(l, m \in \mathbb{Z}, l \geq 0, -l \leq m \leq l, x \in \mathbb{S}^2)$ を基底関数として，

$$\hat{f}^l = \int_{\mathbb{S}^2} f(x) \overline{Y^l(x)} \mathrm{d}x \,, \quad (5)$$

と表わせる．同様に $\mathbb{SO}(3)$ 信号 $f : \mathbb{SO}(3) \to \mathbb{R}^K$ に対する GFT は，球面調和関数が拡張された Wigner の D 行列 $D_{mn}^l(Q)(l, m, n \in \mathbb{Z}, l \geq 0, -l \leq m, n \leq l, Q \in \mathbb{SO}(3))$ を基底関数として，

$$\hat{f}^l = \int_{\mathbb{SO}(3)} f(Q) \overline{D^l(Q)} \mathrm{d}Q \,, \quad (6)$$

と表せる．$Y_m^l(x), D_{mn}^l(Q)$ の直交性により，ψ, f の畳み込みは周波数領域では $\widehat{\psi \star f}^l = \hat{f}^l \cdot \hat{\psi}^{l\dagger}$，すなわち $\hat{f}^l$ と $\hat{\psi}^l$ の随伴行列との行列積で計算できる．

球面 ConvNet は，(3) 式，(4) 式の定義を用いた複数の畳み込みで構築される．各畳み込み層のあとには，正規化層や活性化層などを適宜用いる．入力は $\mathbb{S}^2$ 信号であり，入力層では $\mathbb{S}^2$ 信号の畳み込みを，以降の層では $\mathbb{SO}(3)$ 信号の畳み込みを行う．入出力の $\mathbb{S}^2$ 信号は天頂角 $\beta \in [0, \pi]$ と方位角 $\alpha \in [0, 2\pi]$ に対して，中間表現の $\mathbb{SO}(3)$ 信号は Z-Y-Z オイラー角 $\alpha \in [0, 2\pi], \beta \in [0, \pi], \gamma \in [0, 2\pi]$ に対する信号で表し，各次元に対して同じ長さになるよう離散化して表現する．出力層のあとに Z-Y-Z オイラー角の γ 方向に平均を取ることで，$\mathbb{S}^2$ 信号を得る．

3.5 反射マップ生成ネットワーク

観測反射マップは欠損を含んだ不完全なものであるので，観測反射マップによって条件付けされた GAN を用いて完全な反射マップを復元する．条件付き GAN として，生成器の入力を条件とする pix2pix[9] のフレームワークを用いる．ネットワークは**図 4** に示すように，反射マップを生成する生成器と，生成された反射マップの本物らしさを確か

© 2022 Information Processing Society of Japan

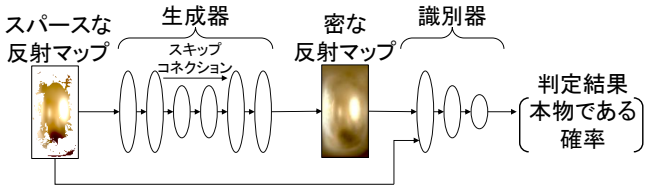

図 4 反射マップ生成ネットワークの構造．観測反射マップから生成器が反射マップを生成し，もとの観測反射マップと反射マップの組を入力として識別器は反射マップの真贋を判定する．

める識別器から構成される．生成器と識別器が，互いに競合しながら学習を進めることで，最終的に生成器が自然な反射マップを生成するように促す．

生成器および識別器のネットワークには，球面 ConvNet を用いる．生成器はエンコーダー・デコーダーで構成され，エンコーダーとデコーダーの間には，スキップコネクションを設ける．識別器はエンコーダーのみからなり，入力される観測反射マップと反射マップの組合せが自然であるかどうかを，本物である確率（$[0,1]$）として予測する．学習時には真の反射マップと観測反射マップの組み合わせに対しても判定を行い学習する．各畳み込み層の後に活性化関数として，エンコーダーでは Leaky ReLU 関数を，デコーダーでは ReLU 関数を挿入する．反射マップの定義域は，$\mathbb{S}^2$ 空間全体のうち視点に写りうる手前の半球のみしかない．球面 ConvNet は全球に対して定義しているので，ネットワークの入出力では反射マップを反対側にも転写して扱う．

生成器と識別器の学習には，真の完全な反射マップに対する損失と，識別器の正誤判定に対する損失を用いる．GAN の目的関数は生成器を G，識別器を D とすると，

$$\mathcal{L}_{\mathrm{GAN}}(G,D) = \mathbb{E}_{x,y}[\log D(x,y)] + \mathbb{E}_{x,y}[\log(1 - D(x,G(x))], \tag{7}$$

と表せる．ここで x, y はそれぞれ観測反射マップ，真の反射マップである．生成器は $\mathcal{L}_{\mathrm{GAN}}$ を最小化するよう，識別器は $\mathcal{L}_{\mathrm{GAN}}$ を最大化するように学習を行う．また，生成器には識別器を間違えさせるだけでなく，真値に近い値を出力する必要があるので，出力に対して真値との L1 損失 $\mathcal{L}_{\mathrm{L1}}(G) = \mathbb{E}_{x,y}[||y - G(x)||_1]$ も与える．したがって求めたい生成器のパラメーター G^* は，

$$G^* = \arg\min_G \max_D \mathcal{L}_{\mathrm{GAN}}(G,D) + \lambda \mathcal{L}_{\mathrm{L1}}(G), \tag{8}$$

で表される．(8) 式に基づき，生成器と識別器を交互に最適化する．

3.6 光源状況推定ネットワーク

レンダリング関数の逆関数を球面 ConvNet を用いて実現し，完全な反射マップを元に光源状況の推定を行う．光源状況推定ネットワークは，反射マップ生成ネットワークによって出力された反射マップを入力として，推定した光源状況を出力する．反射マップは反射マップ生成ネットワークと同様に，半球に対する信号である反射マップを，反対側の半球に転写して扱う．

本研究では光源状況を，鏡面球を写真に撮ったような，鏡面物体に対する反射マップ（以下，鏡面反射マップ）として推定する．一般的にコンピュータグラフィックスなどでは，光源状況を，図 2 の中に例で示したような環境マップで表すことが多い．この表現は，物体の位置を中心として周りを見回したように光源状況をとらえ，天頂角・方位角で光源を表している．反射マップは引数が法線の半球であるのに対し，環境マップは引数が入射方向の全球である．したがって，環境マップを出力の光源状況に用いると，ネットワークはこの定義域やドメインの変化をも学習する必要が生じる．一方で鏡面反射マップは，光源状況を反射マップの空間に合わせた表現方法になっているため，この変化を学習する必要がない．また，鏡面反射の場合入射角と出射角がおなじであることから，反射マップの法線方向に対して光の入射方向は一意に定まるので，鏡面反射マップは直接環境マップに変換することができる．

ネットワークの構造は，基本的に反射マップ生成ネットワークの生成器と同じく，スキップコネクションを持つエンコーダーとデコーダーで構成される．各球面畳み込み層の後には Batch Normalization と活性化関数として ReLU 層を挿入する．また，ネットワークの学習は，出力する鏡面反射マップと真の鏡面反射マップとの平均二乗誤差を用いて行う．

4. 評価実験

本節では，光源状況推定における球面 ConvNet の有用性を検証するために行った，具体的な評価実験の内容とその結果について述べる．

4.1 学習と評価手法

ネットワークの学習に用いるデータセットには，数多くの光源状況と様々な反射特性や物体形状による合成画像を利用した．光源状況として Laval Indoor HDR Dataset[10] を，反射特性には MERL BRDF Database[11] を用い，反射マップと，各光源状況に対応する鏡面反射マップを合成した．Laval Indoor HDR Dataset の光源状況 2233 個は，学習用と評価用にそれぞれ 1787 個，446 個に分割した．また MERL BRDF Database の 100 個の反射特性を，学習用・評価用にそれぞれ 80 個，20 個に分割した．訓練・評価に用いる反射マップはそれぞれ，訓練用・評価用に分けた光源状況および反射特性のみを用いて作成した．光源推定ネットワークは，訓練用の反射マップと対応する鏡面反射マップで学習を行った．既存手法との比較には DeLight-Net[3] のデータセットを用いて評価を行った．

作成した合成反射マップを利用し，反射マップ生成ネッ

© 2022 Information Processing Society of Japan

トワークに用いるデータセットの作成も行った．まず，ランダムな形状の物体 [12] の法線マップをもとに，3.3 節で説明した手法で生成される観測反射マップの，観測できた領域のマスクを作成した．このマスクを，合成反射マップに掛けることで，擬似的に欠損した反射マップと完全な反射マップの真値を持つデータセットを作成した．ランダムな形状の物体の法線マップは 2685 個の各物体について 10 視点で観測したものを用意した．物体についても訓練用と評価用の組にわけ，合成反射マップデータには対応する組からランダムに物体・視点を選択し割り当てた．反射マップ生成ネットワークの訓練と評価にはこのデータセットを用いた．

光源状況はダイナミックレンジの広い情報であるので，物体の画像や反射マップ，光源状況には HDR 画像を用いた．球面畳み込みに限らず畳み込みニューラルネットワークでは，HDR 画像のように値の範囲が広いと，値が大きい高輝度の部分に大きく影響されてしまう．そこで，実際の画像やマップ x を，$\log_{10}(x+\varepsilon)$ の対数スケールに変換した．さらに各ネットワークで扱う値が ConvNet で扱いやすい範囲におおよそ入るよう，入力となるデータセット全体の $\log_{10}(x+\varepsilon)$ の 1, 99 パーセンタイル（それぞれ p_1, p_{99} とする）を求め，その範囲が $[0,1]$ となるよう標準化を行った．

生成された反射マップや推定光源状況を定量的・定性的に比較することで，モデルの有効性を評価した．定量評価の指標には対数平均二乗平方根誤差（LRSME）と DSSIM を用いた．LRSME は対数 $\log(x+1)$ を取った値の L2 距離の二乗の平均の平方根で，画像の情報としての近さを測った．また DSSIM で HDR 画像をトーンマッピングした低ダイナミックレンジの画像の構造的類似度を測った．これらの指標はどちらも，小さい方がより真値に近いことを意味する．

4.2 反射マップ生成ネットワークの精度評価

完全な反射マップ推定の手法について，敵対的生成ネットワーク（GAN）の有効性の評価を行った．提案手法である，(8) 式を用いて生成器と識別器で交互に学習を進めたものに対して，生成器のみを用いて (8) 式のうち真値に対する L1 誤差（第二項）のみを用いて学習を行ったものを比較した．L1 誤差に対して，識別器による誤差（(8) 式第一項）を敵対的誤差と呼ぶ．提案手法および比較手法の生成器のネットワークは同一なものを用いた．また入力として与える不完全な観測反射マップは，観測された領域について 4.1 節で述べた標準化を行った後，欠損した領域は有効領域の平均色で埋めた．

図 5 に定性的比較を示す．L1 損失のみで識別器を用いずに行った場合は，識別器による損失も加えた場合よりも，環境がよく映り込んでいる反射マップにおいて，光源などが若干ぼやけた生成結果となっている．このことは，汎用的な画像変換に対して GAN を用いた pix2pix の論文 [9] で検証された L1 損失と敵対的損失の関係に即している．すなわち，L1 損失は真値との近さに対する制約を低周波成分について与え，敵対的損失はより高周波な鮮明さという制約を与えている．一方で，欠けた部分の補間という点では，同等程度の結果となった．これは GAN の学習の不安定さも影響していると考えられ，GAN の学習が十分ではなかった可能性もある．どちらの手法でも欠けた部分に対して補間が働いており，球面畳み込みが生成モデルにも応用できることがわかった．

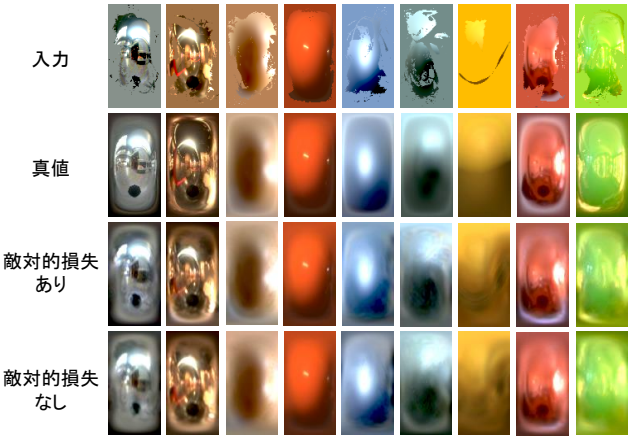

図 5 敵対的損失の影響．敵対的損失があるほうが，若干鮮明な反射マップが復元されている．一方で欠けた部分の復元という点では同等の結果となったが，どちらの手法でも光源などが復元されていることがわかる．

4.3 光源状況推定精度の評価

4.3.1 球面畳み込みと二次元畳み込みの比較

一つの反射特性について，光源状況推定ネットワークによるレンダリング関数の逆問題に対する表現力を評価した．すなわち，反射特性を既知と仮定し，様々な光源状況下に置かれた場合の反射マップと鏡面反射マップにより訓練を行い，同じ反射特性で未知の光源状況下のもとで評価を行った．提案手法の球面 ConvNet に対する比較実験として，この実験を球面畳み込みではなく二次元平面畳み込みでも行った．本提案手法と同様に，スキップコネクションを持ったエンコーダー・デコーダーからなり，各層では平面畳み込みを用いた．また内部パラメーター数については，提案手法の球面 ConvNet におおよそ一致するよう，各層におけるフィルター数を調整した．球面 ConvNet が持つ内部パラメーターが約 90 万個であるのに対し，比較するネットワークは約 100 万個の内部パラメーターを持つ．

図 6, **表 1** に，それぞれ定性的，定量的な結果を示す．定性的な比較では，二次元畳み込みよりも球面畳み込みの方が，写る照明や物体の縁や角がややはっきりと表れている．二次元平面畳み込みでは，フィルターの形状による縦

© 2022 Information Processing Society of Japan

図 6 一つの反射特性に対する球面畳み込みと二次元平面畳み込みの定性比較．球面畳み込みには特異なアーティファクトが出ているが，二次元平面畳み込みよりも鮮明な推定ができている．

表 1 一つの反射特性に対する球面畳み込みと二次元平面畳み込みの定量比較．球面畳み込みではより高い精度を実現できている．

手法	LRMSE	DSSIM
球面畳み込み	0.125	0.081
二次元畳み込み	0.235	0.106

横の粗いアーティファクトが出ている．これによって解像感が球面畳み込みよりも低くなると考えられる．一方で，球面畳み込みでは畳み込み実行時に有限の周波数領域で計算を行ってしまうため，矩形波を有限の係数でフーリエ級数展開したときと同じように，強い光源の周りに独特なアーティファクトが生じている．球面畳み込みの方がアーティファクトは顕著に感じられるが，球面上に置いては二次元平面のアーティファクトより自然なものであり，定量的な評価でも球面畳み込みがより高い精度を上げている．以上から，球面畳み込みはアーティファクトがでてきてしまうものの，二次元平面畳み込みよりも鮮明な推定ができることがわかった．

4.3.2 実データへの適用

用意した合成データで訓練したモデル全体を，DeLight-Net[3] の実データセットに対して評価した．このデータセットは五つの環境下でそれぞれ 40 種類の素材の球状物体と鏡面球を撮影したものである．このうち，Chen[6] らが DeLight-Net との比較に用いた 91 個の組み合わせを，本手法の評価でも用いた．球状の物体の画像に対し，真球と仮定して法線マップを作成し，提案手法を適用後，推定された鏡面反射マップを今度は逆に球の画像に変換した．

図 7 に定性的な結果を示す．既存研究に対して，物体自身の色が完全に分離しきれず推定された光源状況に残ってしまう結果となった．この違いは反射特性の推定を明示的に同時に行うかという違いではないかと考える．既存研究では，光源状況だけでなく反射特性も同時に推定を行う．Chen らの手法では推定された光源状況と反射特性によって物体画像が再現されるようにそれぞれを最適化していく．また Georgoulis らの手法 [3] では，一部を共有した二つの二次元平面畳み込みニューラルネットワークを用いて反射特性と光源状況を推定している．事前学習ではどちらも，反射特性の真値を利用している．本手法では，一つのネットワークで光源状況のみを推定し，事前学習としても光源状況の真値のみを与えて学習させた．このような違いから，既存手法では物体の色が主として反射特性に依存するような推定ができたのではないかと考察する．また，本手法の学習に用いた光源状況は，室内のみで，暖色系の光源が大半であったことも，うまく分離できない理由ではないかと考えられる．

一方で，Georgoulis らの手法と比較して，例えば左端の結果における壁と床の境界や机などのように，細部な部分がより明確に推定されている．また Chen らの手法では，高周波のノイズのような不自然な模様が多くみられるが，提案手法ではあまり見られない結果となった．

表 2 は定量的比較を示すものである．定量評価の計算領域などの具体的な計算条件の詳細が既存研究において不明であり，論文中の値を再現することができなかったため，結果をもとに再計算した．また，光源状況には定数倍の曖昧さがあるので，Chen との比較では各推定結果について LRMSE を最小化する定数を計算し，推定結果にかけた場合も計算した．Georgoulis らの推定結果は低ダイナミックレンジの画像しか得られなかったため DSSIM のみ計算し，彼らの手法に合わせて提案手法の出力を縮小した場合も計算した．表 2 に示すように，LRMSE が倍率調整を行わない場合 Chen らの手法と比べて悪かったが，倍率調整を行った場合精度が良い結果となった．Chen らの手法は，ノイズのような模様がみられるように外れ値が非常に多い推定結果であるからと考えられる．その意味で，提案手法はより自然な推定になっており，DSSIM でも高い精度を出している．また解像度を合わせた比較では，Georgoulis らの手法よりも DSSIM が良い結果となった．

Georgoulis らの手法では光源状況推定のネットワークの畳み込みの重みだけで約 230 万個の内部パラメーターを学習しているのに対し，本手法では図 1 のすべてを合わせても約 178 万個の内部パラメーターのみで表現される．さらに Chen らの手法は推定の際に最適化が必要であるが，本手法では入力から畳み込み計算するだけである．以上のことから，球面畳み込みを用いることで，自然な光源状況推定をコンパクトに実現できることがわかった．

5. 結論

本研究では，球面畳み込みを用いた形状既知な物体の画

© 2022 Information Processing Society of Japan

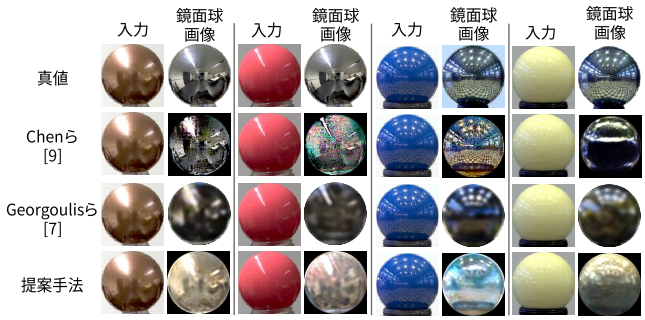

図 7 実データに対する光源状況推定の定性的結果．反射特性の色については分離しきれていないが，Chen らの手法よりノイズが少なく，Georgoulis らの手法より鮮明に推定できている．

表 2 光源状況推定ネットワークの実データに対する定量比較．倍率調整をした場合 Chen らの手法よりも二つの指標で高い精度が出ている．解像度を合わせた比較では Georgoulis らよりも低い DSSIM 誤差を達成した．

手法	LRMSE	DSSIM
提案手法	0.861	0.299
Chen ら [6]	0.832	0.312
提案手法（倍率調整あり）	0.471	0.299
Chen ら [6]（倍率調整あり）	0.601	0.312
提案手法（64x64 解像度）	0.834	0.255
Georgoulis ら [3]（64x64 解像度）	-	0.275

像の光源状況を推定する手法を提案し，球面畳み込みの有効性を示すため，いくつかの評価を行った．球面畳み込みを GAN に用いた評価では，球面畳み込みが生成モデルにも応用できることを示した．また球面畳み込みと，二次元平面畳み込みを比較した評価では，球面畳み込みの周波数成分に対する能力がわかった一方で，アーティファクトが出てしまうこともわかった．実データに対して適用した例では，物体と光源との間の色の分離という問題があったが，定量的比較では高い精度を示し，推論時には最適化を計算する必要なく少ないパラメーター数で光源状況の推定が行えることがわかった．我々は，球面畳み込みが，光源状況推定に有効であることを示すとともに，生成モデルに応用し，球面畳み込みの可能性を広げた．

今後の課題としては，まず，より実際のデータに汎化させ，物体の色を正しく分離することが挙げられる．これにはデータセットの改善も必要で，室内の暖色系の多い光源状況データセットだけを用いるのではなく，より多様な室外や寒色系の光源状況を学習させることで，より正確な光源状況の推定ができるのではないかと考える．また本手法は一つのネットワークで反射マップから光源状況を推定するものであったが，既存研究のように反射特性も明示的に推定できるようにすることでより正確な光源状況推定が実現できる可能性がある．

本研究をさらに発展させた形として，光源状況推定のネットワークを反射特性に応じて変化するものすることで，他研究との融合も考えられる．本手法では，光源状況推定ネットワークの内部フィルターが反射特性によらず固定されたものである．しかし，レンダリング方程式の逆関数という意味では，フィルターは反射特性に応じて変わるべきである．反射特性依存のネットワークを導入すれば，反射特性や物体の形状を推定する既存研究と融合させることができ，一枚の画像のみから，物体の形状，反射特性，光源状況まで推定できる可能性がある．これが実現できれば，拡張現実や状況認識を行うロボティクスなどへの応用がさらに期待できるものとなる．

謝辞

この研究の一部は JSPS 20H05951, 21H04893, JST JPMJCR20G7, JPMJSP2110 の助成を受けて行ったものです．

参考文献

[1] Lombardi, S. and Nishino, K.: Reflectance and Natural Illumination from a Single Image, *Proc. ECCV*, pp. 582–595 (2012).
[2] Lombardi, S. and Nishino, K.: Reflectance and Illumination Recovery in the Wild, *IEEE TPAMI*, Vol. 38, No. 1, pp. 129–141 (2016).
[3] Georgoulis, S., Rematas, K., Ritschel, T., Gavves, E., Fritz, M., Van Gool, L. and Tuytelaars, T.: Reflectance and Natural Illumination from Single-Material Specular Objects Using Deep Learning, *IEEE TPAMI*, Vol. 40, No. 8, pp. 1932–1947 (2018).
[4] LeGendre, C., Ma, W.-C., Fyffe, G., Flynn, J., Charbonnel, L., Busch, J. and Debevec, P.: DeepLight: Learning Illumination for Unconstrained Mobile Mixed Reality, *Proc. CVPR* (2019).
[5] Ulyanov, D., Vedaldi, A. and Lempitsky, V.: Deep image prior, *Proc. CVPR*, pp. 9446–9454 (2018).
[6] Chen, Z., Nobuhara, S. and Nishino, K.: Invertible Neural BRDF for Object Inverse Rendering, *IEEE TPAMI*, No. 01, pp. 1–1 (2021).
[7] Kajiya, J. T.: The Rendering Equation, *Proc. SIGGRAPH*, pp. 143–150 (1986).
[8] Cohen, T. S., Geiger, M., Köhler, J. and Welling, M.: Spherical CNNs, *Proc. ICLR* (2018).
[9] Isola, P., Zhu, J.-Y., Zhou, T. and Efros, A. A.: Image-to-Image Translation with Conditional Adversarial Networks, *Proc. CVPR*, pp. 5967–5976 (2017).
[10] Gardner, M.-A., Sunkavalli, K., Yumer, E., Shen, X., Gambaretto, E., Gagné, C. and Lalonde, J.-F.: Learning to Predict Indoor Illumination from a Single Image, *ACM TOG*, Vol. 36, No. 6 (2017).
[11] Matusik, W., Pfister, H., Brand, M. and McMillan, L.: A Data-Driven Reflectance Model, *ACM TOG*, Vol. 22, No. 3, pp. 759–769 (2003).
[12] Xu, Z., Sunkavalli, K., Hadap, S. and Ramamoorthi, R.: Deep image-based relighting from optimal sparse samples, *ACM TOG*, Vol. 37, No. 4, p. 126 (2018).

© 2022 Information Processing Society of Japan