Vol.2020-CVIM-222 No.16
2020/5/15

Conditional GANによる 双方向テクスチャ関数の圧縮と補間

山本 悠一郎[1]　日浦 慎作[1]

概要：素材表面の視覚的質感を再現する表現力の高い手段として双方向テクスチャ関数 (Bidirectional Texture Function: BTF) が用いられるが，BTF の獲得には様々な照明，観測方向の組み合わせについてその都度サンプルの計測が必要になる．例えば照明，観測方向の解像度を 5 度刻みとすると約 168 万回の計測が必要な膨大なものとなる．また，表面座標の解像度によってはデータセットが巨大になる空間的コストも問題となる．そこで本研究では，これらのコスト削減のために Conditional GAN (CGAN) を用い，照明方位・観測方向の角分解能に関する補間とそのデータ圧縮法について検討した．

1. はじめに

視覚的質感の再現には写真をただ一枚撮影するだけでは不足する場合が多い．物体は光源の方向や物体を観測する角度によって見え方が変化し，この見え方の違いが視覚的質感に大きく影響している．これらの見え方の変化は，表面の微細な凹凸や各点の反射特性の違いによる複雑な現象であるから，物理現象に基づくモデルでは表現に限界がある．そこで，物理モデルに基づかない質感再現の手段として双方向テクスチャ関数 (Bidirectional Texture Function: BTF) が用いられる．BTF は照明方向, 観測方向, 表面座標の 6 自由度を入力，画素値を出力とする仮想的な関数である．原理上，BTF は均一な照明下における見え方の変化のすべてを表現できる．

様々な照明，観測方向の組み合わせについてその都度サンプルの撮影を行い，先の 6 自由度に対する画素値を記録することで，現実世界の素材についての BTF の近似が得られる．しかし，BTF を獲得するための撮影には長時間がかかる．例えば照明, 観測方向の解像度を 5 度刻みとすると約 168 万回の計測が必要となり, これは 1 秒に 1 回の高速な計測であっても約 19.4 日かかる膨大なものとなる．後述する UBO2003 Datesets[1] を公開している Bonn 大学では図 1 のような，カメラと光源をドーム状に配置した設備を有している．つまり，実用可能な BTF を獲得するには図 1 のような大掛かりな設備を使用するか，一台のカメラ・光源装置で長時間かけて計測を行うかをしなければならない．さらに，BTF の表面座標の解像度によっては獲得

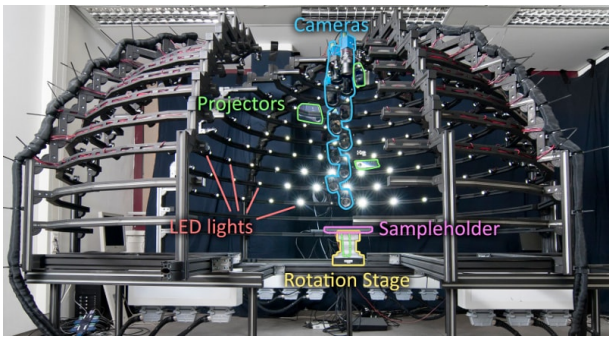


図 1: Bonn 大学の BTF 計測装置 [2]

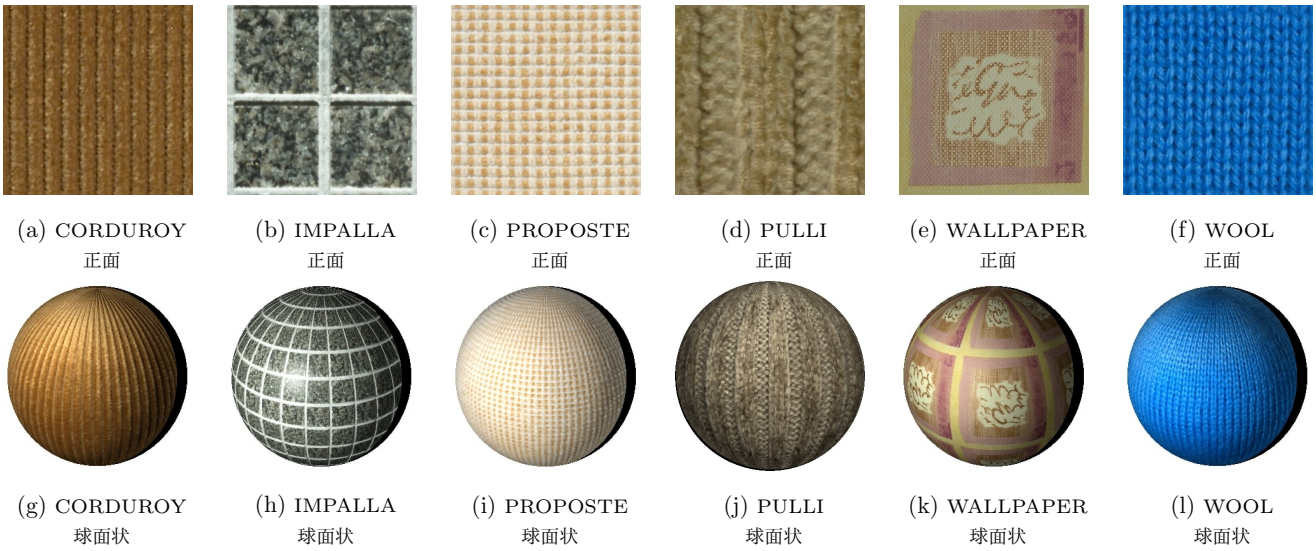
したBTF のデータセットが巨大になる空間的コストも問題となる．168 万枚の 512x512，24bit/画素のフルカラー画像では無圧縮で約 10TB の容量を使用する．よって，質感再現の手段として BTF を気軽には採用できない．そこで本研究では，GAN を用いて照明・観測方向の角分解能に関する補間方法の検討を通し，BTF 獲得のための撮影時間の短縮やデータセットの圧縮方法について検討する．

2. 関連研究 - 機械学習による BTF の生成

Zhang らは BTF データの生成に深層学習を用いる方法を検討し，Conditional variational AutoEncoder (CVAE) と比較して SSIM による比較では Conditional GAN の方が優れていたことを報告している [3].

この論文では，Mirko らの UBO2003 Datasets[1] を BTF データセットとして用いている．このデータセットは 6 つの素材について，照明 81 方向，観測 81 方向，表面座標 256x256 の解像度の計測で BTF を取得している．図 2 に UBO2003 Datasets のサンプルを示す．図 2(a) から図 2(f)

[1] 兵庫県立大学



(a) CORDUROY 正面　(b) IMPALLA 正面　(c) PROPOSTE 正面　(d) PULLI 正面　(e) WALLPAPER 正面　(f) WOOL 正面

(g) CORDUROY 球面状　(h) IMPALLA 球面状　(i) PROPOSTE 球面状　(j) PULLI 球面状　(k) WALLPAPER 球面状　(l) WOOL 球面状

図 2: UBO2003 Datasets[1] のサンプル

が実際にデータセットに含まれる画像であり，図 2(g) から図 2(l) はデータセットを公開している Bonn 大学がこのデータセットを元に，素材を球面に貼り付けた時の見た目をレンダリングしたものである．

Zhang らの論文は, この内 CORDUROY, PULLI, WOOL を用いて実験している．アルゴリズムの評価は，データセットに含まれる 6,561 の計測条件における画像を再生成し，生成した画像 1 枚ずつについてデータセット画像との SSIM を算出し，その平均値により行っている．

Zhang らが用いた GAN の構造は, 照明方向, 観測方向の 4 パラメーターに加え，3 つの素材を同一のモデルで学習させたため素材の種類を示す 1 パラメーターをラベル情報とし，これと 200 次元の潜在変数を Generator モデルの入力に用いている．Discriminator モデルへのラベル情報の入力は中間層に行っている．

表 1 に CVAE と Conditional GAN との場合の評価結果の比較を示す．CVAE と比較して Conditional GAN による生成画像のほうがデータセットをよりよく再現できていることが分かる．図 3 に実際の生成画像を示す．Conditional GAN による生成画像のほうが，特に色味についてよく再現されていることが分かる．

表 1: アルゴリズムの違いによる SSIM 値の比較 [3]

	CORDUROY	PULLI	WOOL
CVAE	0.2078	0.1052	0.2645
Conditional GAN	0.3104	0.3805	0.4281

3. 実験と結果

実験は，2 節で紹介した UBO2003 Datasets[1] をデータセットとして用いた．この実験では試行回数を稼ぐため 64x64 の解像度に中央をクロップしたものを学習に用いている．

図 3: アルゴリズムの違いによる生成画像の違い
各行は上から，CVAE による生成画像，Conditional GAN による生成画像，正解画像
各列は左から，CORDUROY，PULLI，WOOL

3.1 直交座標系による生成条件の連続化

照明方向と観測方向について，データセットでは天頂角と方位角で管理されている．この形式を便宜上極座標と呼ぶことにする．ある角度について教師データを極座標による 2 パラメーターで与えると，方位角が連続的に取り扱われない場合が発生する．例えば，方位角が 1 度と 359 度との差は実質的には 2 度しかないが，このパラメーターをそのまま扱うと 358 度の差として取り扱われる．

そこで，この境界が連続的に取り扱われるように，極座標を直交座標に変換したものを教師データとして用いることにした．天頂角を θ[deg]，方位角を ϕ[deg] とすると変換後の x，y は，

$$x = \frac{\theta}{75} \times \cos\phi \quad (1)$$

$$y = \frac{\theta}{75} \times \sin\phi \quad (2)$$

となる．ただし，75deg はデータセットに含まれる最大天頂角である．

また，この変換後は $x^2 + y^2 \leq 1$ となるので，学習時に Generator モデルに入れる教師データ用の変数は，この範囲内をランダムに一様に分布するようにする．

図 4 に，実験で得た生成画像と正解画像との比較を示す．図 4 から生成画像の品質が向上していることがわかる．

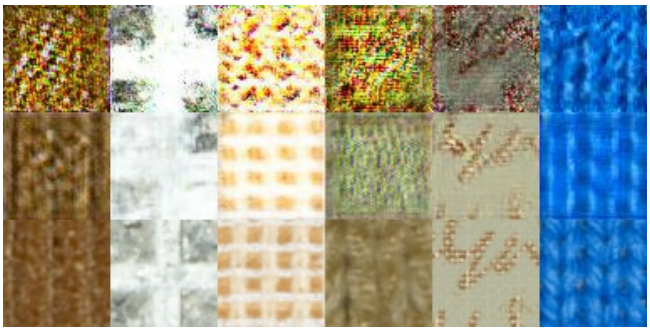


図 4: 教師データの与え方による生成画像 (観測方向・光源方向とも真上) の違い
各行は上から，極座標を用いた場合の生成画像，直交座標を用いた場合の生成画像，正解画像
各列は左から，CORDUROY，IMPALLA，PROPOSTE，PULLI，WALLPAPER，WOOL

3.2 潜在変数の除去

通常の Conditional GAN では，同一の教師データに対しても複数の画像が生成されることを期待する．例えば，数字の書き分けでは「4」という数字にも複数の書き方がある．そこで教師データとは別に Generator モデルの入力に潜在変数を置けば，同じ教師データでも潜在変数が異なることで様々な画像を生成することが出来る．しかし，BTF では同一の教師データに対して同じ画像が生成されても構わないため，潜在変数の有無によって品質に大きな差がない場合は，潜在変数を使わないことで学習難易度の低減とモデルサイズの削減を期待できる．

図 5 に実験で得た生成画像と正解画像との比較を示す．

また，SSIM と PSNR による評価を表 2 に示す．ただし，潜在変数の選び方で生成画像が変化するため，3 回画像を生成し，SSIM・PSNR はそれら平均を用いている．

図 5 で実際の生成画像を比較すると，IMPALLA は潜在変数を使用すると SSIM や PSNR による評価では悪くなっていたが，画像を見るとコントラストが上がっており，視感上では潜在変数を用いた方良くなっているようにも見える．逆に，WALLPAPER では潜在変数を用いた方が SSIM

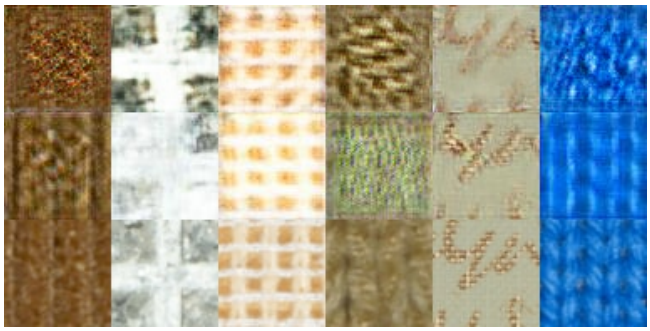


図 5: 教師データの与え方による生成画像 (観測方向・光源方向とも真上) の違い
各行は上から，潜在変数 100 次元を使用した生成画像，潜在変数を使用しなかった生成画像，正解画像
各列は左から，CORDUROY，IMPALLA，PROPOSTE，PULLI，WALLPAPER，WOOL

表 2: 潜在変数の有無による生成画像と正解画像の誤差

(a) 潜在変数を用いなかった場合

素材の種類	SSIM	PSNR (dB)
CORDUROY	0.296	19.7
IMPALLA	0.388	18.0
PROPOSTE	0.306	18.4
PULLI	0.163	19.3
WALLPAPER	0.340	20.6
WOOL	0.273	19.7

(b) 100 次元の潜在変数を用いた場合

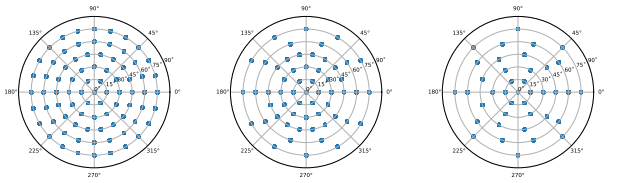
素材の種類	SSIM	PSNR (dB)
CORDUROY	0.338	19.7
IMPALLA	0.350	16.8
PROPOSTE	0.292	18.1
PULLI	0.159	18.9
WALLPAPER	0.388	21.0
WOOL	0.264	19.6

や PSNR による評価では良くなっていたが，画像を見るとぼやけた印象があり，視感上では潜在変数を用いないほうが良いように見える．学習にはランダム性があるため，この結果では潜在変数はあってもなくても生成画像の品質には大きな影響がないように思える．よって提案手法では潜在変数は用いないこととした．

3.3 正解画像 (データセット) の規模を減らした学習

今回の実験では 6,561 枚の画像を BTF の学習に使用したが，実用を考えた場合，もっと少ないデータセットの規模でもある程度の品質が確保できるのかどうかが課題となる．今回はデータセットを半分にした場合，4 分の 1 にした場合で実験を行った．データセットを間引く際は，人間が手動でなるべく均等になるように配置したものと完全

にランダムに間引いたものの 2 通りを実験している．図 6 に半分，4 分の 1 に手動で間引いた後の照明・観測方向を示す．



(a) 全て (81 方向, 6,561 枚) (b) 半分 (57 方向, 3,249 枚) (c) 4 分の 1(40 方向, 1,600 枚)

図 6: 手動によるデータセットの間引き方

データセットの規模を半分・4 分の 1 にしたとき，エポック数を 2 倍・4 倍にすれば重み更新の回数を揃えられる．データセット全てを用いて 100 エポック学習したときを基準にし，重み更新の回数を揃えてデータセットの規模を変えながら学習させて得た生成画像に対し，SSIM，PSNR を計算した結果を表 3 に示す．ただし，SSIM，PSNR は学習

表 3: データセット規模の違いによる SSIM 値と PSNR 値

データセット	間引き方	エポック数	SSIM	PSNR (dB)
全て	-	100	0.388	18.0
半分	手動	200	0.372	17.7
半分	ランダム	200	0.367	17.3
4 分の 1	手動	400	0.385	18.1
4 分の 1	ランダム	400	0.391	18.0

に用いなかったデータについても算出を行っている．間引き方を比較すると大差がない事がわかる．無作為抽出は，均等に間引いたときに近付いていくこともあり，手動で間引くことによって良い結果を得ようとする場合は，疎密をつけた方が良いことが示唆される．また，データセットの規模によらず同じような SSIM，PSNR が得られていることも分かる．

手動で間引いた半分・4 分の 1 のデータセットで，データセットに含まれる画像と含まれない画像を分けて生成画像と比較したものを表 4 に示す．正解画像に対する誤差は，

表 4: 学習に含まれる画像と含まない画像に対する誤差

データセットの規模	比較対象画像	SSIM	PSNR (dB)
半分	学習時に含まれる	0.406	18.2
半分	学習時に含まない	0.338	17.2
4 分の 1	学習時に含まれる	0.429	18.7
4 分の 1	学習時に含まない	0.371	17.8

学習時に用いた画像の方が小さくなっていることが分かるが，学習時に用いなかった画像も大きな誤差とはなっていない．

データセットから間引かれた画像について生成した画像を比較したものを図 7 に示す．表 4 と図 7 より，学習時に

図 7: データセットに含まれなかった画像の生成結果
「観測方向は真上，光源方向は天頂角 45 度・方位角 340 度」の条件で画像を生成
各列は左から，正解画像，学習にデータセット全体を使用，学習に半分のデータセットを使用，学習に 4 分の 1 のデータセットを使用

用いなかった画像についても尤もらしく画像が生成されていることが分かる．

この結果から，大量のデータセットを用意して学習させる以外に，学習上重要な画像を用意して学習時間をかけることでも生成画像の精度が上がることが示唆される．

4. まとめ

本研究では，BTF の補間に Conditional GAN を用いる方法を提案した．学習に用いなかった正解画像に対しても尤もらしい生成画像が得られたことから，BTF を獲得するために計測した条件以外の素材表面の画像を Conditional GAN で連続的に補間できることが確認できた．ただし，実際に計測で得られた画像と Conditional GAN により生成した画像とが完全には一致しない問題はある．機械学習アルゴリズムを使う以上，正解画像と生成画像を完全に一致させるのは極めて困難だが，教師データの変換をし結果が向上したことを踏まえ，モデルの学習方法を見直すことで両者の誤差を小さく出来る余地があることも感じられた．

今回の研究では，データセットの効率的な削減方法を発見するには至らなかったが，小さい規模のデータセットでも生成画像の精度を上げられる余地があることが分かった．

参考文献

[1] Sattler, M., Sarlette, R. and Klein, R.: ”Efficient and realistic visualization of cloth”, *Rendering Techniques*, pp. 167–178 (2003).

[2] Schwartz, C., Sarlette, R., Weinmann, M. and Klein, R.: ”DOME II: A Parallelized BTF Acquisition System”, *Eurographics Workshop on Material Appearance Modeling: Issues and Acquisition* (Rushmeier, H. and Klein, R., eds.), Eurographics Association, pp. 25–31 (online), DOI: 10.2312/MAM.MAM2013.025-031 (2013).

[3] Zhang, X., Dong, J., Gan, Y., Yu, H. and Qi, L.: ”BTF data Generation based on Deep Learning”, *Procedia computer science*, Vol. 147, pp. 233–239 (2019).