折り紙の折り方を説明する動画における手の除去

山中　颯人†　　岡部　誠†

静岡大学†

1. 概要

折り紙は子供から大人まで楽しめる紙工作の手法であり，折り方を説明する動画は YouTube などの動画共有サイトにおいて人気のコンテンツである．しかし，これらの動画を参考に自分で折り紙を折ろうとすると，説明者の手によって隠れる部分が多いためにどのように紙を動かせばよいのかが理解できない場面に遭遇することがある．そこで，我々は折り紙の説明動画から説明者の手を除去し，紙の動きのみが見えるような動画を合成することを目的とした．本稿では 2 つのアプローチを用いて実験を行った．１つ目はエッジを用いた手法(EdgeConnect[1])である．画像内の欠損部分を含むエッジを推定するエッジ推定器と、そのエッジをもとに RGB 画像を生成する画像生成器の 2 つを利用し画像補完を行う手法である．我々はエッジ推定器として U-Net を使用し，学習を行うためのデータセットは CG を用いて作成した．画像生成器は既存のものを使用した．２つ目は拡散モデル[2]を用いた手法である．拡散モデルとはノイズからノイズ除去を繰り返し行うことで画像を生成するものであり，拡散モデルを利用した画像補完手法である RePaint[3]の動画への応用を実験した．以上の２つのアプローチを使い実際の折り紙の画像に対してどの程度手を除去できるのかを既存手法の結果と比較した．

2. アプローチ

一般的な画像補完では構造を修復することが難しい(図１)．そこで我々はエッジを用いた手法と拡散モデルを用いた手法の 2 種類のアプローチを用いて実験を行う．手を消して画像補完を行うため手の部分が画像の欠損部分として考える．

図 1．左）入力画像　右）画像補完の結果

2.1 エッジを用いた手法

我々は U-Net を用いてエッジ推定器を新たに作成し，EdgeConnect[1]のフローを参考に実験を行った．

2.1.1 EdgeConnect

EdgeConnect とは「人が絵を描くときまず下書きをしてから色を塗る」という考え方にインスピレーションを受け画像補完にエッジを利用した手法である．入力画像内の欠損部分のエッジを推定するエッジ推定器と，推定したエッジから RGB 画像を作り出す画像生成器の 2 工程で行われる(図 2)．エッジを使うことで従来の手法では

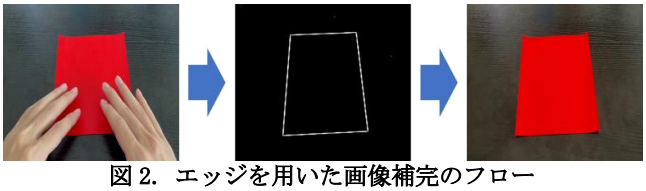

図 2．エッジを用いた画像補完のフロー

苦手であった構造修復を得意としている．

2.1.2 エッジ推定器の作成

我々はエッジ推定器として U-Net を使用した．入力は RGB 画像，欠損部分のマスク，欠損部分以外のエッジの 3 つを使用し，出力は欠損部分を含む全体のエッジである．学習するためのデータセットとして手が写っていない折り紙のみのエッジが必要になる．折っている途中の形を大量に集めるのは非常に困難であるので，CG を用いてデータセットを作成した．紙のみが動き鶴が出来上がる３D アニメーションを作成し，角度を変えながらレンダリングすることで画像のデータセットを約 10 万枚作成した(図 3)．

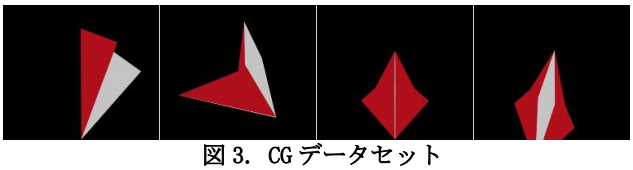

図 3．CG データセット

2.2 拡散モデルを用いた手法

拡散モデルを用いて画像補完を行う手法である RePaint[3]の動画への応用を実験した．

2.2.1 拡散モデルと画像補完

拡散モデルとは標準正規分布に従うランダムなノイズからノイズ除去を繰り返すことで画像を生成するモデルである．RGB 画像にノイズを加えていく逆の過程を学習させることでノイズ除去が可能になった．

拡散モデルを画像補完に応用させたのが RePaint である．RePaint は欠損部分をノイズから徐々に生成していく(図 4)．そのノイズ除去の過程で既知の領域の情報を欠損部分に入れながらノイズ除去をすることで図 4 の一番右の画像のように人間が見ても違和感のない画像が生成できる．

図 4．拡散モデルを用いた画像補完

2.2.2 動画への応用

折り紙の動画に適用する場合 1 フレーム前は現在のフレームと非常に似た折り紙の形であるため，1 フレーム前の画像を使って欠損部分を補完するのが有用であると考えたが，1 フレーム前の画像が完全に補完できていない場合，次のフレームに支障がでてしまう．そこで今回

Removal of hands in a video explaining how to fold origami

† Hayato Yamanaka, Makoto Okabe, Graduate School of Engineering, Shizuoka University

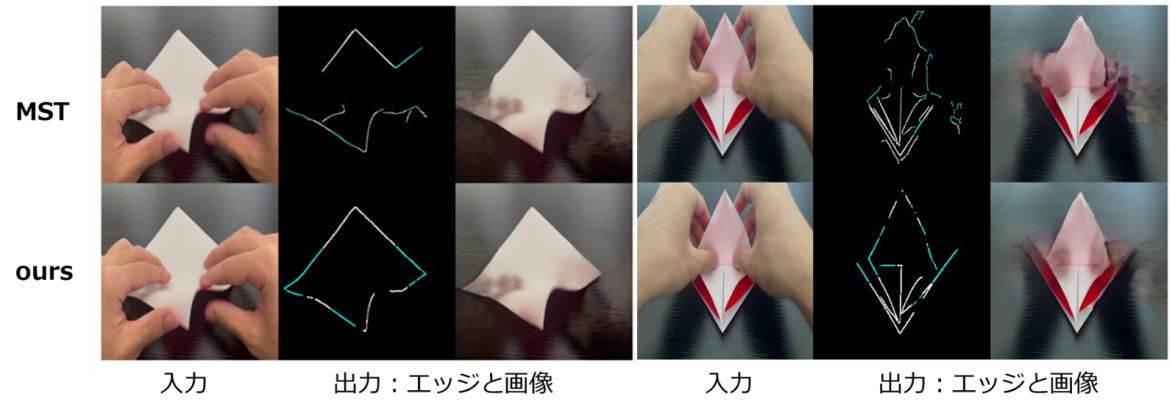

図 5. エッジを用いた手法による結果

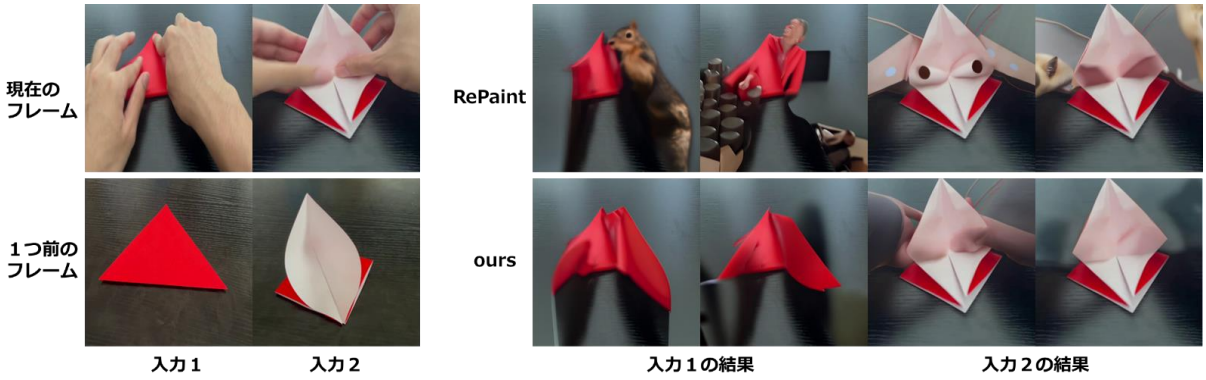

図 6. 拡散モデルを用いた手法による結果（出力結果にはランダム性がある）

は 1 フレーム前の画像の代わりに，現在のフレームと似ている折り紙の形の画像を使用した．RePaint では欠損部分に完全なノイズを使用しているが，我々は完全なノイズではなく，1 フレーム前の画像にノイズを加えたものを使用した．

3. 結果と考察

3.1 エッジを用いた手法

エッジを用いた手法の出力結果を図 5 に示す．図 5 は上段が既存手法である MST(Multi-scale Sketch Tensor inpainting) [4]の結果，下段が新しく作成したエッジ推定器を用いた結果であり，それぞれ 2 つの場面の同じ入力画像に対しての結果となっている．

まず図 5 の左側の結果では MST は手で隠れている折り紙の形が推定出来ていないのに対して，本手法では見えていない部分に対してもきれいなエッジが生成できているのがわかる．その後，推定したエッジを画像生成器に入力し画像を生成すると，MST ではぼやけて形がはっきりわからないが，本手法はくっきりとした形が確認できる．一方で図 5 の右側の結果では MST も本手法も手で隠れている部分に対して完璧なエッジを推定することはできず，最終的な画像もはっきりとした折り紙の形は確認できなかった．

本稿では良い結果と悪い結果の２つを示したが，１つの動画を通しては悪い結果の割合が高く動画としてきれいなものは作成できなかった．今回画像 1 つ 1 つに対して独立で処理をしていたので動画としての応用が必要になってくるだろう．

3.2 拡散モデルを用いた手法

拡散モデルを用いた手法の結果を図 6 に示す．右側の上段が既存手法である RePaint の結果，右側の下段が本手法の結果である．RePaint の入力は現在のフレーム（図 6 左側の上段）のみの画像 1 枚を使い，欠損部分は標準正規分布に従うノイズを使用し画像補完を行う．一方，本手法の入力は現在のフレームと 1 つ前のフレーム（図 6 左側の上下）の２つの画像を使用し，現在のフレーム内の欠損部分に対して 1 つ前のフレームから作るノイズを用いて画像補完を行う．拡散モデルは確率過程であるので同じ入力から様々な結果が得られる．

既存手法の RePaint では画像補完する際に全く関係ないものや動物を生成してしまっているのに対して，本手法では生成されている部分は折り紙に近いものであることがわかる．本手法の入力 2 の結果では，人が想像する手のない画像に近いものが生成で来た．一方で入力 1 の本手法の結果は形が悪いものとなった．

拡散モデルは学習済みのものを使用したため折り紙と関係ないものも生成されてしまったので新たに折り紙を学習させることで形の復元がより良いものになると考えられる．またこの手法の前提条件として 1 フレーム前の画像が正確に復元されていないといけない．この前提条件を緩和する手法についても検討していきたい．

4. 参考文献

[1] K.Nazeri, E.Ng, T.Joseph, F.Z.Qureshi, M.Ebrahimi, “EdgeConnect: Structure Guided Image Inpainting using Edge Prediction”, In ICCV2019

[2] J.Ho, A.Jain, P.Abbeel, “Denoising Diffusion Probabilistic Models”, In 2020

[3] A.Lugmayr, M.Danelljan, A.Romero, F.Yu, R.Timofte, L.V.Gool, “RePaint: Inpainting using Denoising Diffusion Probabilistic Models”, In CVPR 2022

[4] C.Cao, Y.Fu, “Learning a Sketch Tensor Space for Image Inpaiting of Man-made Scenes”, In ICCV2021