Journal of the Korean Society of Surveying, Geodesy, Photogrammetry and Cartography
Vol. 40, No. 1, 23-30, 2022
https://doi.org/10.7848/ksgpc.2022.40.1.23

ISSN 1598-4850(Print)
ISSN 2288-260X(Online)
Original article

실감모형 제작을 위한 고해상도 유물 이미지 매팅

High Resolution Photo Matting for Construction of Photo-realistic Model

최석근[1] · 이승기[2] · 최도연[3] · 김광호[4]
Choi, Seok-Keun · Lee, Soung-Ki · Choi, Do-Yeon · Kim, Gwang-Ho

Abstract

Recently, there are various studies underway on the deep learning-used image matting methods. Even in the field of photogrammetry, a process of extracting information about relics from images photographed is essential to produce a high-quality realistic model. Such a process requires a great deal of time and manpower, so chroma-key has been used for extraction so far. This method is low in accuracy of sub-classification, however, it is difficult to apply the existing method to high-quality realistic models.

Thus, this study attempted to remove background information from high-resolution relic images by using prior background information and trained learning data and evaluate both qualitative and quantitative results of the relic images extracted. As a result, this proposed method with FBA(manual trimap) showed quantitatively better results, and even in the qualitative evaluation, it was high in accuracy of classification around relics. Accordingly, this study confirmed the applicability of the proposed method in the indoor relic photography since it showed high accuracy and fast processing speed by acquiring prior background information when classifying high-resolution relic images.

Keywords : Deep Learning, Artifact Image, Photo-realistic Model, Matting, Pre-background Information

초 록

최근 딥러닝을 이용한 이미지 매팅 방법에 관한 다양한 연구가 진행되고 있다. 특히, 사진측량 분야에서도 고품질의 실감모형을 제작하기 위해서는 촬영된 이미지에서 유물 정보를 추출하는 과정이 필요하며, 이와 같은 과정은 많은 시간과 인력이 들어 기존에는 크로마키를 이용하여 추출하는 방법이 많이 활용되고 있다. 그러나, 기존의 방법은 세부 분류에 대한 정확도가 떨어져 고품질 실감모형에 적용하기에는 어려움이 있었다. 본 연구에서는 사전배경정보와 훈련된 학습데이터를 이용하여 고해상도 유물 이미지에서 배경정보를 제거하고 추출된 유물 이미지에 대하여 정성적, 정량적 결과를 평가하였다. 그 결과 제안된 방법과 FBA(매뉴얼 트라이맵)이 정량적으로 높은 결과를 나타냈으며, 정성적 평가에서도 유물 주변부의 분류도가 높은 정확도를 보였다. 따라서 제안된 방법은 고해상도 유물 이미지 분류에 있어 사전배경정보 취득을 통하여 높은 정확도와 빠른 처리 속도를 나타냈으며, 실내 유물 촬영에서 그 활용 가능성을 확인하였다.

핵심어 : 딥러닝, 유물 이미지, 실감모형, 매팅, 사전배경정보

Received 2022. 01. 13, Revised 2022. 01. 27, Accepted 2022. 02. 04

1) Member, Dept. of Civil Engineering, Chungbuk National University(E-mail : skchoi@cbnu.ac.kr)
2) Member, Terrapix research fellow(E-mail: iruril@chungbuk.ac.kr)
3) Terrapix research fellow(E-mail: jedgoh@naver.com)
4) Corresponding Author, Member, Dept. of Civil Engineering, Chungbuk National University(E-mail : kgh7759@naver.com)

This is an Open Access article distributed under the terms of the Creative Commons Attribution Non-Commercial License (http://creativecommons.org/licenses/by-nc/3.0) which permits unrestricted non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited.

1. 서 론

이미지 매팅은 전경(foreground) 물체와 배경(background)을 분류하여 배경을 교체하거나 전경만을 추출하여 활용하게 하는 방법이다. 최근에는 매팅을 활용하여 영화의 특수효과, Zoom, Google Meet, Microsoft Teams과 같은 화상회의에서도 많이 사용되고 있다. 사진측량 분야에서도 고품질의 실감 모형을 제작하기 위해서는 촬영된 이미지에서 유물의 정보를 추출하는 과정이 필요하며, 이와 같은 과정은 많은 시간과 인력이 들어 기존에는 크로마키를 이용하여 추출하는 방법이 많이 활용되고 있다. 그러나, 이와 같은 크로마키는 조명 및 각도 등의 제약이 많으며, 세부 추출 결과 또한 좋지 않아 이미지 매팅을 이용한 연구가 많이 수행되고 있다. 그러나, 기존의 이미지 매팅 방법은 고해상도 이미지을 가지는 유물의 전경을 추출하는데 있어 몇 가지 제약을 가지고 있다. 크로마키 기법과 마찬가지로 기존의 매팅방법은 전체적인 배경을 추출하는 데는 효과적이지만, 미세한 세부 사항(헤어, 내부 경계 등)에서 명확하지 않은 추출결과물을 생성하고, 트라이 맵, 스크래처와 같은 복잡한 사용자 정보 입력과 적용 해상도의 제한과 같은 한계점이 있다(Chuang *et al*.,2001; Sun *et al*.,2004; Levin *et al*.,2007; Gastal *et al*.,2010; Lin *et al*.,2021). 따라서, 본 연구에서는 복잡한 사용자 정보를 배제하고 고해상도(2,000만화소 이상) 이미지에 적용하기 위하여 사전 배경정보 매팅방법을 유물 이미지에 적용하고 활용 가능성을 검증하였다.

최근 들어, 이미지 매팅에서 가장 많이 다뤄지는 문제는 크게 분할(segmentation)과 매팅방법론(분류를 위한 입력데이터형태) 등 2가지로 구분된다. 분할은 영상의 각 화소가 어떠한 속성을 가지는지를 할당하는 기법이며 매팅에서 가장 많이 활용되는 방법은 Mask RCNN과 DeepLabV3+이다. Mask RCNN은 COCO 데이터 세트을 기반으로 한 기존의 분류자에서 높은 성능을 보여주고 있는 분할방법이다(He *et al*.,2017). DeepLabV3+은 Chen *et al*.(2018)이 제안한 최신의 분할알고리즘이며, Ni *et al*.(2021)은 DeepLabV3+를 이용하여 고해상도 이미지의 분할을 수행하였다.

매팅방법은 크게 3가지의 방법으로 나누어지며, 첫 번째 트라이맵을 이용하는 방법 두 번째 자동분할 방법을 이용한 방법 마지막으로 배경정보를 이용한 방법이 있다. 대표적인 트라이맵을 이용한 연구로 Xu *et al*.(2018)은 트라이맵과 매팅 데이터세트을 딥러닝에 적용하여 매팅을 수행하였으며, Liu *et al*.(2020) 는 트라이맵의 품질에 따른 영향을 최소화하기 위하여 트라이맵의 노이즈를 고려하여 매팅을 수행하였다. 트라이맵과 같은 사용자 정보 없이 매팅을 수행하기 위하여 Xiaoyong *et al*.(2016)는 CNN (Convolutional Neural Network)을 이용하여 초상화의 매팅을 수행하였고, Bingke *et al*.(2017)은 딥러닝과 모바일 폰 이미지를 이용하여 사람 이미지를 실시간으로 매팅하였다. Qiao *et al*.(2020)은 HAttMatting (Hierarchical Attention Matting Network)을 이용하여 외부정보없이 사람 이미지에 대한 매팅을 성공적으로 추출하였으나, Sengupta *et al*.(2020)의 연구에서는 사전정보 매팅추출에 비하여 정확도가 낮은 것으로 나타났다. 사전배경 정보추출 매팅은 기존의 Bayesian과 Poisson 매팅 방법에서는 트라이맵이 필요하였으나, Sengupta *et al*.(2020)은 BGM (BackGround Matting)에서 사전배경정보를 통하여 매팅을 수행하였다. 하지만 BGM의 경우 입력데이터의 한계가 있어 고해상도 이미지에는 적용하기에 한계가 있다(Lin *et al*.,2021).

최근의 연구들은 대부분 Adobe사의 인물 데이터베이스를 기반으로 훈련을 수행하고 있으며, 매팅의 분류 목적도 인물에 대한 배경 분류로 이루어져 있어 유물 매팅에 바로 적용이 불가능하며, 유물에 대한 트라이맵 생성에 대한 연구 또한 미비하다. 또한 2,000만 화소 이상의 카메라에서 취득되는 이미지에 대한 분류 또한 제한적이므로 기존의 BGM과 같은 방법에 적용하여 그 가능성을 확인하기 어려운 문제가 있다.

이를 위하여, 본 연구에서는 Lin *et al*.(2021)의 사전배경정보를 이용한 매팅방법을 적용하여 고해상도 유물 이미지에 대한 매팅을 수행하였다. Lin *et al*.(2021)의 제안 방법은 인물에 대한 분류 알고리즘이지만 훈련데이터의 추가를 통하여 이종의 물체에 대한 분류를 수행할 수 있으며, 입력데이터의 크기가 비교적 다른 방법에 비하여 자유롭고 트라이맵과 같은 복잡한 사전정보가 필요하지 않아 유물 이미지 분류에 있어 적용성이 크므로 본 연구에 적용하였다. 유물 분류 훈련을 위하여 다양한 형태의 매팅 데이터인 AIM-500, AISegment.com에서 제공하는 Matting Human Datasets과 연구자가 생산한 다양한 유물의 촬영 이미지와 알파이미지를 적용하였고, 실험을 통해 고해상도 유물 이미지에 대하여 매팅 성능과 적용성을 평가하였다.

2. 연구방법

2.1 이미지 매팅을 위한 딥러닝 모델

매팅을 위해서는 새로운 배경(B')과 새로운 이미지(I')가 생성되어야 하며, 이를 위해서는 입력 이미지(I)와 배경정보(B)정보가 필요하다. 이미지 (I)와 배경정보(B)가 주어지면

Eq. (1)에 의하여 알파 매트(α)와 전경(F)를 예측하여 새로운 배경(B')에 의한 새로운 이미지 (I')가 생성된다.

$$I' = \alpha F + (1-\alpha)B' \quad (1)$$

where, α is alpha matte, F means the foreground, B' is the new background, and I' is the new image to composite over any new background.

고해상도 입력 이미지에 저해상도 전경 잔차를 적용할 수 있게 하여 고해상도 이미지의 매팅을 적용하기 위해서는 전경을 직접적으로 계산하는 대신에 Eq. (2)를 이용하여 전경 잔차(F^R)를 계산한다.

$$F^R = F - I \quad (2)$$

where, F^R is foreground residual, and I is the given an image.

계산된 전경 잔차는 이미지(I)에 전경 잔차(F^R)를 Eq. (3)을 이용 추가하고 전경(F)값을 복원하게 된다.

$$F = \max(\min(F^R + I, 1), 0) \quad (3)$$

딥러닝을 이용한 매팅에서 단일 네트워크를 이용한 고해상도 이미지 매팅은 메모리 소비와 계산 속도에 있어서 많은 한계점이 발생하므로, 본 연구에서는 두 개의 네트워크를 적용하여 메모리 소비와 계산 속도의 향상을 높였다. 두 개의 네트워크는 기본 네트워크인 G_B와 리파인 네트워크 인 G_R으로 구성되어 있으며, Fig. 1과 같이 Lin *et al.*(2021)의 구성 네트워크를 기반으로 입력데이터와 파라메터 값을 수정하여 유물 이미지 매팅에 활용하였다.

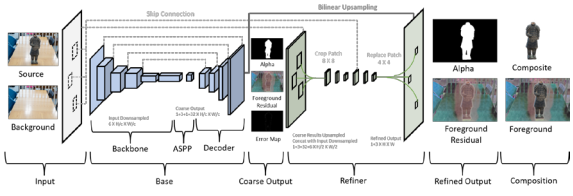


Fig. 1. The base network G_B operates on the downsampled input to produce coarse-grained results and an error prediction map. The refinement network G_R selects error-prone patches and refines them to the full resolution (Lin *et al.* 2021)

2.1.1 기본 네트워크(G_B)

기본 네트워크의 형태는 DeepLabV3+의 형식을 따르며, 세부 구성은 백본, ASPP 및 디코더로 구분된다. 백본은 ResNet-101을 활용하고, ASPP 모듈은 3, 6 및 9의 다중 컨볼루션 필터로 구성되며, 디코더는 각 단계에서 업샘플링을 적용하여 배치 정규화, ReLU 활성화 과정을 수행한다.

기본 네트워크에서는 1차 알파 매트, 전경 잔차(저해상도), 에러 예측 값, 히든 피처 값이 추출되며, 히든 피처 값은 리파인 네트워크에 사용될 전역 값을 포함한다.

2.1.2 리파인 네트워크(G_R)

리파인 네트워크에서 기본 네트워크에서 생성된 1차 알파매트와 전경잔차를 이용하여 입력 이미지(고해상도)의 매트 값을 복구한다. 기본 네트워크가 전체 이미지에서 작동하는 반면, 리파인 네트워크는 오류 예측 데이터에 따라 선택된 패치에서만 작동되고, 처음에는 원래 해상도의 절반의 해상도에서 시작하며, 이후 전체 해상도에서 작업을 수행한다. 리파인 네트워크에서 사용되는 패치의 개수(k)는 사전에 설정된 값을 기반으로 하거나 계산 시간과 결과 값을 고려하여 조정한다.

기본 네트워크에서 예측된 오류 데이터는 다운 샘플링된 데이터로 4분에 1로 데이터가 다운 샘플링 되었다면, 오류 맵의 하나의 픽셀은 원래 해상도의 4*4 패치에 해당되며, 리파인 네트워크에서 계산될 오류값이 높은 상위 패치의 개수(k)를 선택하여 계산된다. 따라서, 원래 해상도에서 계산될 픽셀의 총 수는 16×k값이 된다.

Fig. 1과 같이 두 단계의 리파인 과정을 수행하게 되며, 일단계로 1차 알파 매트 값, 전경잔차, 히든 피처, 입력 이미지 및 배경 등을 원래 해상도의 절반으로 재심플링하고, 각각의 특징을 연결한다. 이후 오류 데이터의 상위 패치 값 주변을 8×8 패치로 잘라내고, 4×4로 줄이는 패딩, 배치 정규화 및 ReLU 등을 거친다. 입력 데이터 (I) 및 배경 (B)에서 추출된 8×8 패치와 연결하기 위하여 패치를 다시 8×8로 업샘플링 과정을 수행하고, 배치 정규화 및 ReLU(마지막 레이어 제외)가 있는 3x3 회선의 두 레이어를 추가로 적용하여 전경 잔차와 알파매트를 계산한다.

최종적으로 알파 매트와 전경 잔차를 원래 해상도로 업샘플링하고 최종 알파 매트와 전경 잔차를 얻기 위해 라파인된 4x4 패치와 교체한다.

2.1.3 손실함수

본 연구에서는 훈련결과 α와 사용자 데이터 α^*의 손실함수 계산을 위하여 전체 알파 매트와 엣지(Sobel)에 대해 $L1$

손실을 사용하였다. α 손실함수 계산식은 Eq.(4)와 같다.

$$L_\alpha = \| \alpha - \alpha^* \|_1 + \| \nabla \alpha - \nabla \alpha^* \|_1 . \quad (4)$$

where, α^* is ground-truth, α means the learning data, L_α is the loss of α, $\nabla \alpha$ means the Sobel gradien of α, and $\nabla \alpha^*$ is the sobel gradien of α^*.

Eq.(3)을 이용하여 전경 레이어를 계산할수 있으며, Eq.(5)와 같이 계산된 전경(F)과 실제 전경(F^*)의 L1 손실(L_F)은 $\alpha^* > 0$, 즉 실제 알파값이 배경인 곳만 계산한다.

$$L_F = \| (\alpha^* > 0) * (F - F^*) \|_1 . \quad (5)$$

where, F^* is ground-truth, F means the learning data, and L_F is the loss of F.

리파인 영역 선택을 위해 실제 에러맵(E^*)을 $E^* = |\alpha - \alpha^*|$로 정의한다. 예측 오차 맵과 실제 오차 맵의 오차 맵의 손실(L_E)은 Eq. (6)과 같이 오차 사이의 평균 제곱 오차를 손실로 계산한다.

$$L_E = \| E - E^* \|_2 . \quad (6)$$

where, E is ground-truth, E^*means the learning data, and L_E is the loss of E.

이 손실함수는 예측된 알파와 실제 알파 간의 차이가 큰 경우 예측된 오류 맵이 더 큰 값을 갖도록 하며, 예측된 알파가 향상됨에 따라 훈련 중 반복을 통해 실제 오류 맵이 변경된다.

기본 네트워크 $(\alpha_c, F_c^R, E_c, H_c) = G_B(I_c, B_c)$는 원래 이미지 해상도의 $\frac{1}{c}$에서 작동하며 Eq. (7)의 손실 함수로 훈련된다.

$$L_B = L_{\alpha_c} + L_{F_c} + L_{E_c} . \quad (7)$$

where, L_{α_c} is the loss of alpha matte($\frac{1}{c}$ of the original image resolution), L_{F_c} means the loss of foreground($\frac{1}{c}$ of the original image resolution), L_{E_c} means the loss of error map($\frac{1}{c}$ of the original image resolution), and L_B is the loss of base network.

리파인 네트워크 $(\alpha_c, F^R) = G_R(\alpha_c, F_c^R, E_c, H_c, I, B)$의 손실함수는 아래 Eq. (8)과 같다.

$$L_R = L_a + L_F . \quad (8)$$

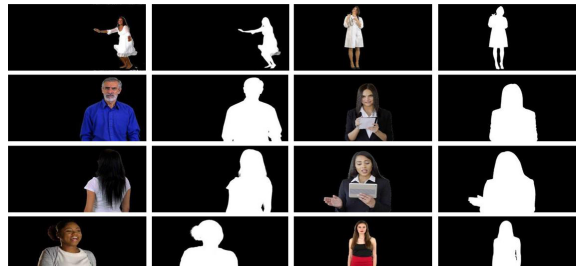
where, L_a is the loss of alpha matte), L_F means the loss of foreground, and L_R is the loss of refine network.

2.2 훈련데이터 세트

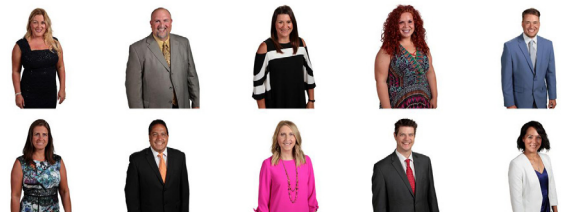
유물 이미지 매팅을 위한 대용량 데이터 세트를 얻는 것은 어렵기 때문에 본 연구에서는 공개 알파 매트 데이터 세트과 연구자가 생성한 유물 데이터 세트을 이용하여 훈련데이터를 구성하였다.

2.2.1 공개 알파 매트 데이터 세트

VideoMatte는 Lin *et al.*(2021)가 제공하고 있는 공개 알파 매트 데이터로 사람에 대한 크로마 키 소프트웨어 Adobe After Effects를 사용하여 484개의 고해상도 그린 스크린 비디오를 수집하고, 총 240,709개의 고유한 알파 매트 및 전경 프레임을 생성한 데이터이다. PhotoMatte는 스튜디오에서 촬영된 크로마키 데이터를 이용하여 동일한 연구자가 수동 조정 및 오류 수정을 통해 추출된 13,665개의 매트 데이터를 제공하고 있다. Fig. 2는 데이터 세트의 2가지 타입의 제공 형태이다.



(a) VideoMatte240K



(b) PhotoMatte13K/85

Fig. 2. Two large-scale matting datasets containing 240k unique frames and 13k unique photos (Lin *et al.*,2021)

SIMD (Semantic Image Matting Dataset) 데이터는 Yanan *et al.*(2021)가 제공하고 있는 공개 알파 매트 데이터로 20개 클래스에 대한 알파 매트 데이터를 제공하고, 복잡한 형태의 매팅 패턴을 제공하고 있는 특징이 있다. Fig.3은 제공되고 있는 복잡한 형태의 알파매트를 보여주고 있다.

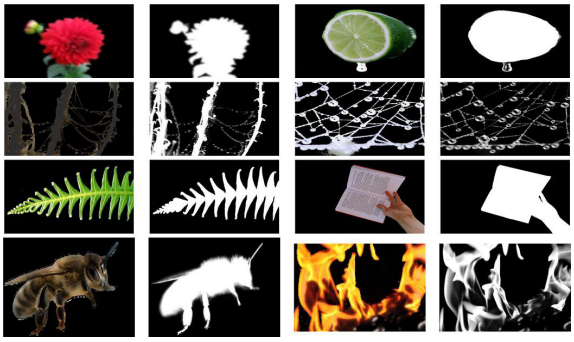


Fig. 3. Semantic image matting dataset containing various types of matting data (Yanan *et al.*,2021)

2.2.2 사용자 알파 매트 데이터 세트

본 연구에서 고해상도 사용자 알파 매트 데이터를 제공하기 위하여 유물 151점에 대한 매팅 데이터를 생성하고 이를 훈련데이터에 추가하였다. Fig.4는 본 연구에 적용된 훈련데이터 형태이다.

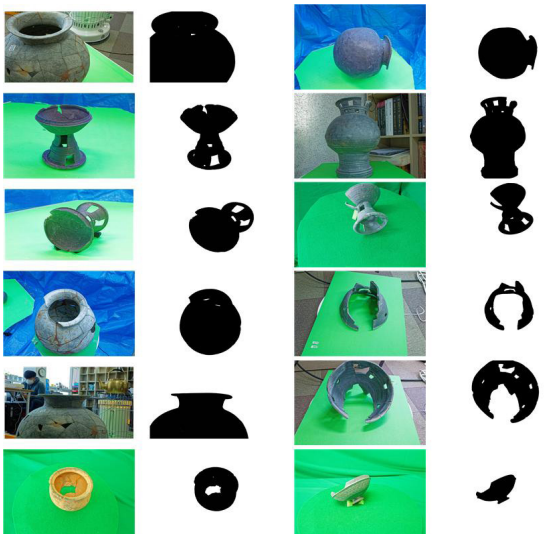


Fig. 4. Artifact image matting data containing 151 types

2.3 학습방법

유물 이미지 매팅 학습은 2.2의 학습자료를 이용하였으며, Python과 Tensorflow를 이용하여 수행하였다. 본 연구에서 학습 및 테스트에 사용한 컴퓨터 제원은 Table 1과 같다. 학습은 G_B의 학습을 선행하고, 이후 세부데이터의 보정 및 추출 결과물 보정을 위하여 G_B와 G_R에 VideoMatte와 PhotoMatte 데이터를 적용하였다. 다음으로 인물이 아닌 다른 물체의 추출 데이터를 얻기 위하여 SIMD을 추가하여 이종 물체의 추출 결과물을 보정하고, 마지막으로 실제 유물 데이터를 추가하여 고해상도 세부 사항을 개선하였다. 특히 실제 유물 데이터는 고해상도(5,060만화소)로 최고 품질이며, 사람이 주석으로 처리한 알파매트가 포함되어 있어 생성된 알파매트의 품질을 개선하는데 매우 유용하다. G_B와 G_R의 훈련에서는 데이터의 정확도를 높이고 입력 데이터를 증가시키기 위하여 매팅된 이미지와 다양한 형태의 배경이미지를 합성하여 훈련을 진행하였다. 각 단계별 학습방법은 Fig.5와 같다.

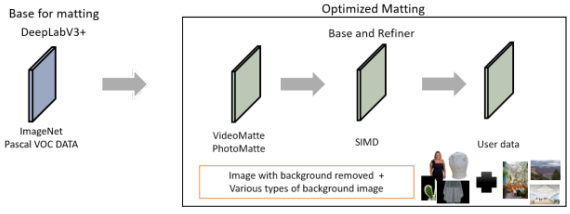


Fig. 5. Flow chart of Matting data training

Table 1. Tested Computer Specification

CPU	Intel(R) Core(TM) i7-9700 CPU @ 3.6.0 GHz
RAM	128G
GPU	NVIDIA GeForce RTX 3090

본 연구에서 사전 훈련된 DeepLabV3+(ImageNet 및 Pascal VOC 데이터) 가중치를 사용하여 기본 네트워크를 훈련하였다. 사용된 옵티마이저는 가장 대표적으로 사용되는 Adam 옵티마이저를 이용하였으며, c=4(다운샘플링 값), k = 20,000(패치 개수)를 사용하였다(Lin *et al.*,2021). 모든 훈련은 수렴할때까지 수행하였으며, 기본 네트워크 훈련은 대해 배치 크기 8과 학습률[1e-4, 5e-4, 5e-4]을 사용하였고, 공동으로 훈련할 때는 배치 크기 4와 학습률[5e-5, 1e-4, 3e-4]을 사용하였다.

3. 실험결과

고해상도 이미지에서 유물 이미지 매팅을 하기 위해 학습자료를 이용하여 학습을 수행하고, 이종 데이터를 이용하

여 고해상도 이미지 매팅 결과를 기존의 매팅 방법인 Forte and Pitié(2020)의 FBA(F,B,Alpha matting)와 Sengupta *et al.*(2020) BGM을 비교평가하였다. 기존 매팅방법은 깃허브 코드를 이용하여 학습하였으며, BGM의 경우 본 연구 훈련 데이터를 적용한 결과값(BGM_{our})을 포함하여 비교·평가하였다.

3.1 유물 데이터 이미지 매팅 정량적 평가

실제 유물 데이터 세트의 테스트 샘플을 5개 추출하여 테스트 훈련데이터를 구성하였다. FBA에 사용될 트라이맵은 사용자가 주석으로 생성하였으며, Rhemann *et al.*(2009)의 이미지 매팅 평가방법인 알파 및 전경에 대한 MSE (Mean Squared Error), SAD (Sum of Absolute Difference), Grad (spatial-gradient metric) 및 Conn (connectivity)를 이용하여 평가하였다.

Table 2는 유물 데이터 세트에 대한 정량적 평가 결과이고, 본 연구의 제안방법과 FBA가 자동 매팅 방법인 BGM에 비하여 SAD에서 9에서 18로 약 2배 차이를 보이고 있으며, MSE에서는 최저 0.06과 최고 0.22로 3배 차이를 보여주고 있다. 또한 Grad에서도 최저 3,800과 최고 28,000로 6.5배과 Conn의 경우는 최저 6,500최고 14,000으로 약 2배 차이를 보여주고 있다. 특히 자동 매팅 방법인 BGM에서는 유물데이터로 훈련한 BGM_{our}에서도 제안방법과 FBA가 높은 결과값을 보여주고 있다. FBA에 비하여 제안 방법이 낮은 결과값(SAD의 경우 0.5, MSE에서는 0.03)을 보여주고 있으나, FBA의 경우 각 이미지에 대하여 사용자 주석 트라이 맵을 이용하여 매팅을 수행하기 때문에 실제 매뉴얼 매팅값과 유사한 분류값을 가지고 있는 것으로 판단되며, 이와 같은 매뉴얼 데이터 입력이 없는 제안 방법이 FBA와 유사한 값을 가지는 것은 매팅 자동화에 있어 효과적인 방법으로 활용될 수 있음을 확인할 수 있었다. 특히, 다수의 이미지를 매팅하는 과정에서 FBA 방법은 각 이미지에 대하여 사용자 트라이 맵을 입력하는 과정이 추가되며, 이에 따라 주석 입력 시간과 사용자에 따른 입력 부정확성에 대한 결과의 신뢰도 저하와 같은 단점이 있다.

Table 2. Quantitative evaluation on different datasets

METHOD	SAD	MSE	Grad	Conn	
FBA	**9.18**	**0.061**	**3,879**	**6,518**	manual trimap
BGM	20.25	0.226	28,384	16,398	
BGM_{our}	18.45	0.201	25,557	14,191	ours dataset
Ours	9.68	0.099	6,536	6,788	

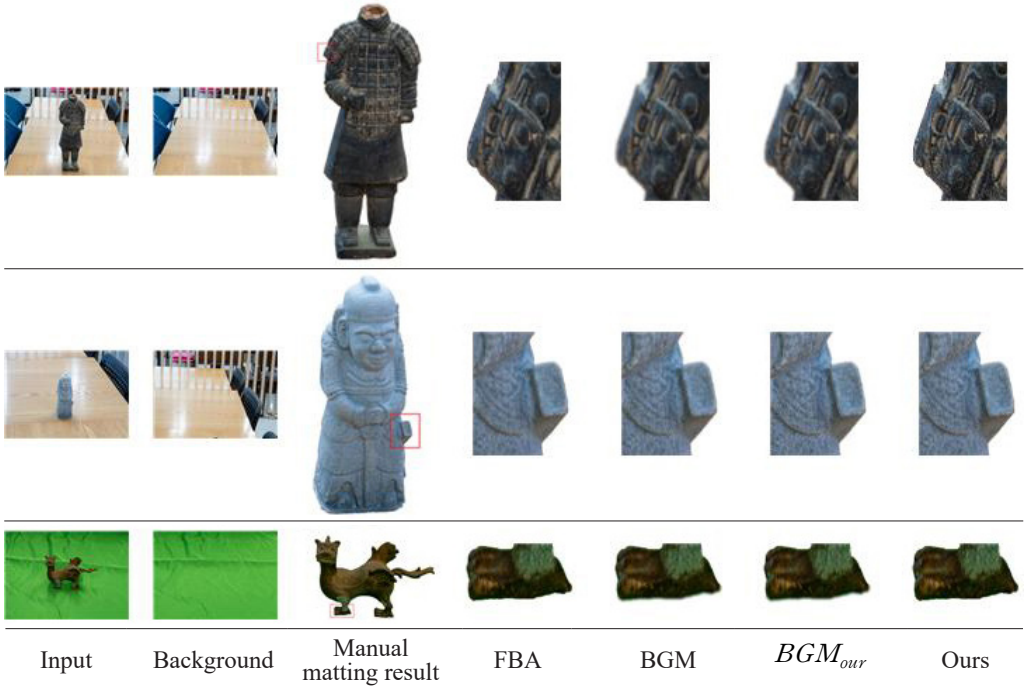


Fig. 6. Results of image matting

3.2 유물 데이터 이미지 매팅 정성적 평가

기존의 매팅방법과 제안된 방법에서 실제 유물 이미지 매팅이 어떻게 추출되는지는 정량적 평가와 별도로 중요하기 때문에 실제 데이터를 평가하기 위하여 고해상도 유물 이미지(5040만 화소)를 이용하여 각 방법별 유물이미지 매팅을 수행하였다. Fig.6은 각 방법별 유물 이미지 매팅 결과이며, 제안된 방법과 매뉴얼 트라이맵 기반의 FBA가 다른 방법에 비하여 유물 가장자리 주변에 대하여 좀더 세밀한 분류를 보여주고 있는 것을 알 수 있다. 또한, 제안 방법은 BGM과 FBA과 달리 전처리 과정없이 원본 해상도를 직접적으로 활용할 수 있어 입력데이터의 한계가 있는 다른 방법에 비하여 활용도가 높으며, 특히 원본 해상도를 유지하면서 세부 분류가 가능하다. 유물 이미지의 경우 카메라 위치 셋팅 후 사전배경 정보 취득을 수행한다면, 제안된 방법이 다른 매팅 방법에 비하여 높은 정확도와 빠른 시간(매뉴얼 정보 입력 불필요)의 유물 이미지 추출이 가능하며, 입력 이미지 한계에 따라 이미지 추출과 병합과 같은 불필요한 전처리 과정 또한 필요가 없어 사전 소요 시간 또한 줄어들어 성능 향상에 영향을 미칠 것으로 판단된다.

4. 결 론

본 연구에서는 기존 이미지 매팅방법의 경우 매뉴얼 데이터를 필요로 하는 단점(비용 및 시간)과 고해상도 이미지 처리에 한계점(입력 데이터 크기 한계)을 극복하기 위하여 사전배경 정보를 활용한 딥러닝 유물 이미지 매팅의 활용성을 분석하였다. 특히, 유물 이미지 매팅의 정확한 추출을 위하여 기존 연구자들의 데이터와 연구자가 취득한 고해상도 유물 이미지를 이용하고 데이터 증대를 위하여 배경정보가 제거된 이미지와 임의의 배경정보를 합성하여 훈련에 사용하였다. 생성된 학습자료를 매팅 모델에 적용하고 고해상도(5,000만 화소)이미지에 대한 성능을 평가하였다. 제안된 방법은 G_B에 대하여 기존 데이터를 이용하여 분류 훈련 후 G_B와 G_R의 통합 훈련은 기존 매팅 데이터를 통하여 정확도를 높였으며, 마지막으로 사용자 데이터를 추가하여 세부정보에 대한 결과값(매뉴얼 훈련인 FBA과 유사한)을 높일 수 있었다. 훈련된 데이터와 사전배경정보를 통하여 제안된 방법은 기존의 이미지별 사용자 정보(트라이맵, 스크래처)를 필요로 하는 것과 달리 여러 장의 고해상도 이미지에 대한 매팅을 빠르고 정확하게 수행할 수 있었다. 또한, 제안된 방법은 1차 오류맵 생성을 통하여 오분류가 발생하는 지역만을 계산하여 중복 계산을 줄이고 고해상도 이미지 매팅에서 분류할 필요가 있는 영역만을 분류하여 효율적인 매팅을 수행할 수 있었다. 하지만, 5,000만 화소의 이미지의 경우 오분류 지역만을 추출하는 경우에도 그래픽카드의 메모리(본 연구의 경우 24GB)의 한계에 의하여 처리 시간이 많이 걸리는 단점이 있었다. 하지만, 기존의 적용의 한계를 극복하고 고해상도 이미지에 적용이 가능하기 때문에 실내에서 턴테이블에서 취득된 유물 이미지의 경우 고정된 카메라 위치와 배경은 동일하기 때문에 제안된 방법을 이용한 이미지 매팅 방법의 효율성이 가장 커지며, 이를 이용한 3차원 실감모형 제작의 활용도 또한 큰 것으로 판단된다.

그러나, 실외 촬영의 경우 사전배경정보를 취득하여야 정확한 매팅이 수행되는 한계점이 있어, 향후 추가적인 연구를 통하여 사전배경정보 없이 단일 이미지 영상만을 이용하여 고해상도 이미지 매팅 방법을 제시할 수 있을 것으로 기대된다.

감사의 글

본 연구는 2019학년도 충북대학교 학술연구지원사업에 의하여 연구되었고, 한국연구재단 (교육부) 기초과학연구프로그램 지원사업에 의해 수행된 연구임(NRF-2018R1D1A1B07048841).

“이 논문은 2020학년도 충북대학교 연구년제 사업의 연구비 지원에 의하여 연구되었음(This work was financially supported by the Research Year of Chungbuk National University in 2020)”

References

Chen, L.C., Zhu, Y., Papandreou, G., Schroff, F. and Adam, H.(2018), Encoder-decoder with atrous separable convolution for semantic image segmentation, *In Proceedings of the European conference on computer vision*, pp. 801-818.

Chuang, Y.Y., Curless, B., Salesin, D.H. and Szeliski, R.(2001, December), A bayesian approach to digital matting, *In Proceedings of the 2001 IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, Vol. 2, pp. II-II.

Forte, M. and Pitié, F.(2020), F, B Alpha Matting, *arXiv preprint arXiv:2003.07711*.

Gastal, E.S. and Oliveira, M.M.(2010, May), Shared sampling for real-time alpha matting, *In Computer Graphics Forum*, Vol. 29, No. 2, pp. 575-584.

He, K., Gkioxari, G., Dollár, P. and Girshick, R.(2017), Mask r-cnn, *In Proceedings of the IEEE international conference*

on computer vision, pp. 2961-2969.

Levin, A., Lischinski, D. and Weiss, Y.(2007), A closed-form solution to natural image matting, *IEEE transactions on pattern analysis and machine intelligence*, Vol. 30, No. 2, pp. 228-242.

Lin, S., Ryabtsev, A., Sengupta, S., Curless, B.L., Seitz, S.M. and Kemelmacher-Shlizerman, I.(2021), Real-time high-resolution background matting, *In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 8762-8771.

Liu, J., Yao, Y., Hou, W., Cui, M., Xie, X., Zhang, C. and Hua, X.S.(2020), Boosting semantic human matting with coarse annotations, *In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 8563-8572.

Ni, J., Zhang, F., Ma, F., Yin, Q. and Xiang, D.(2021), Random region matting for the high-resolution PolSAR image semantic segmentation, *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, Vol. 14, pp 3040-3051.

Qiao, Y., Liu, Y., Yang, X., Zhou, D., Xu, M., Zhang, Q. and Wei, X.(2020), Attention-guided hierarchical structure aggregation for image matting, *In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 13676-13685.

Rhemann, C., Rother, C., Wang, J., Gelautz, M., Kohli, P. and Rott, P.(2009, June), A perceptually motivated online benchmark for image matting, *In 2009 IEEE Conference on Computer Vision and Pattern Recognition*, pp. 1826-1833.

Sengupta, S., Jayaram, V., Curless, B., Seitz, S.M. and Kemelmacher-Shlizerman, I.(2020), Background matting: The world is your green screen, *In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 2291-2300.

Shen, X., Tao, X., Gao, H., Zhou, C. and Jia, J.(2016, October), Deep automatic portrait matting, *In European conference on computer vision*, pp. 92-107.

Sun, J., Jia, J., Tang, C.K. and Shum, H.Y.(2004), Poisson matting, *In ACM SIGGRAPH 2004 Papers*, pp. 315-321.

Sun, Y., Tang, C.K. and Tai, Y.W.(2021), Semantic image matting, *In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 11120-11129.

Xu, N., Price, B., Cohen, S. and Huang, T.(2017), Deep image matting, *In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 2970-2979.

Zhu, B., Chen, Y., Wang, J., Liu, S., Zhang, B. and Tang, M.(2017, October), Fast deep matting for portrait animation on mobile phone, *In Proceedings of the 25th ACM international conference on Multimedia*, pp. 297-305.