# Foreground-aware Image Inpainting

Wei Xiong, Jiahui Yu, Zhe Lin, Jimei Yang, Xin Lu, Connelly Barnes, Jiebo Luo

Iustin Petrisor
Alexandru Sarbu

# Purpose of research

- important problem in computer vision that has many applications
- image editing
- restoration
- composition

# Introduction

- The article focuses on hole filling tasks encountered commonly when removing unwanted regions or objects from photos.
- Filling holes in images with complicated foreground and background composition is one of the most significant and challenging scenarios.
- Conventional inpainting methods typically fill missing pixels by matching and pasting patches based on low level features.

# Introduction

- Recently, deep learning-based methods have emerged as a promising alternative avenue by treating the problem as learning an end-to-end mapping from masked input to completed output.
- However, this has not been an easy task even for state of-the-art models, such as PartialConv and GatedConv.
- In the end, this article proposes a foreground-aware image in painting system that explicitly incorporates the foreground object knowledge into the training process.

# Introduction

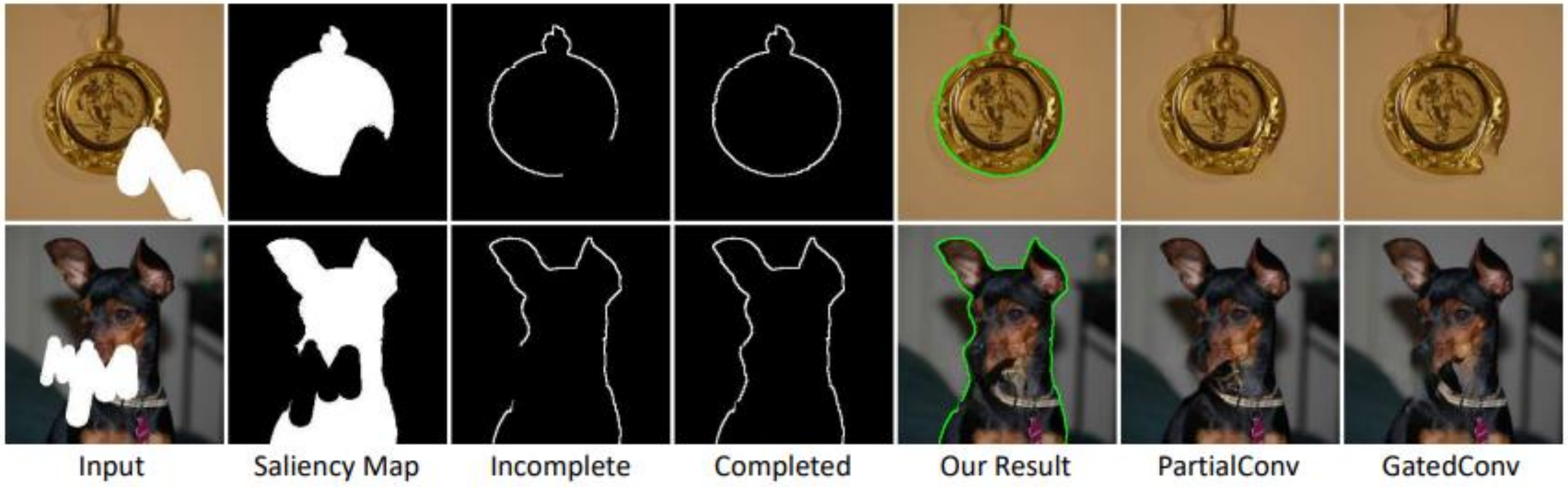
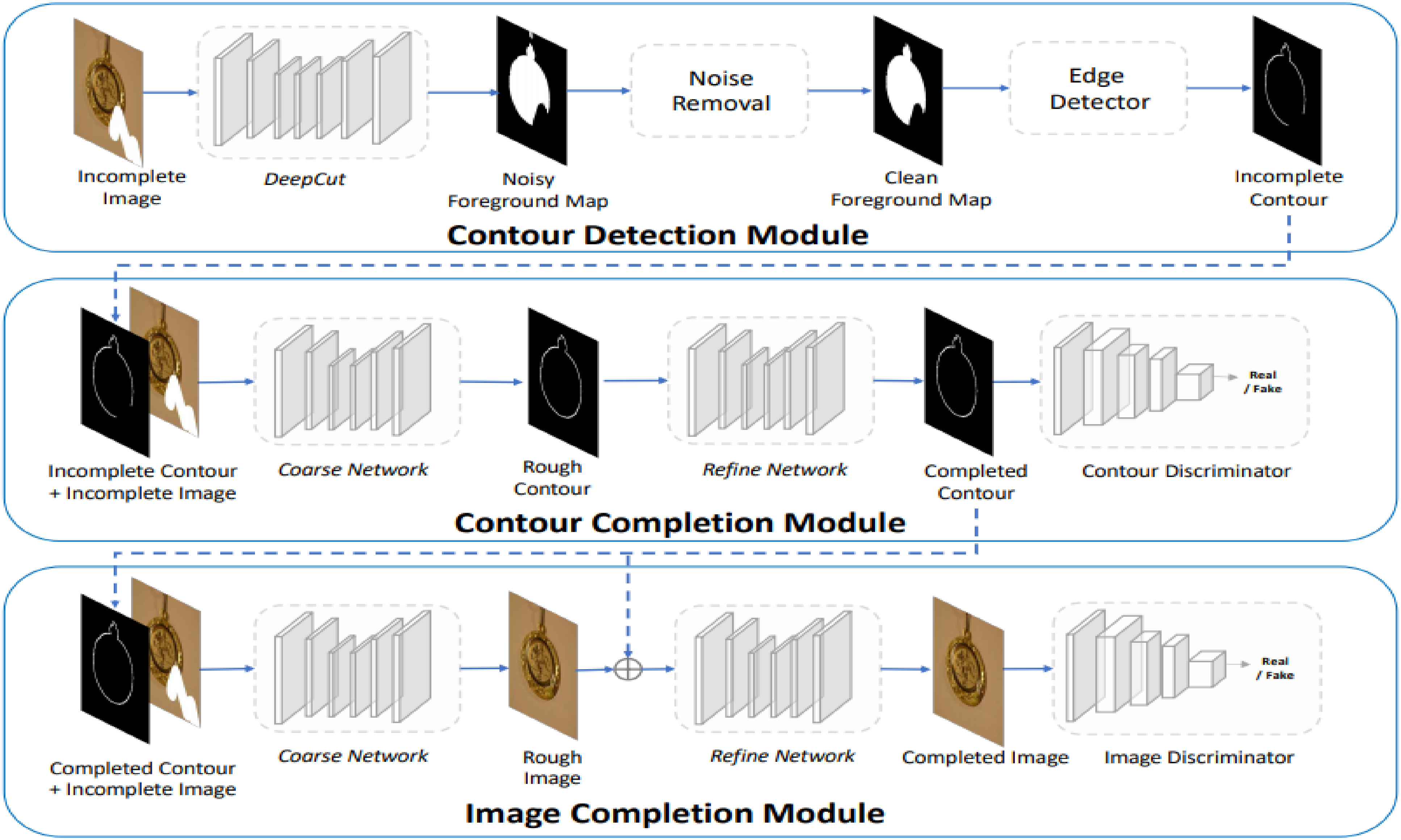



Fig. 1

The disentanglement of structure inference and image completion is conceptually simple and highly effective.

# Approach

- Given an incomplete image, our goal is to output a complete image with a visually pleasing appearance.
- It is a cascade of three modules: incomplete contour detection module, contour completion module and image completion module.



Incomplete Image
DeepCut
Noisy Foreground Map
Noise Removal
Clean Foreground Map
Edge Detector
Incomplete Contour
Contour Detection Module
Incomplete Contour + Incomplete Image
Coarse Network
Rough Contour
Refine Network
Completed Contour
Real / Fake
Contour Discriminator
Contour Completion Module
Completed Contour + Incomplete Image
Coarse Network
Rough Image
Refine Network
Completed Image
Real / Fake
Image Discriminator
Image Completion Module


Fig. 2

# Data Acquisition and Hole Generation

- For training they collected 15,762 natural images that contained one or several salient objects, from a variety of public datasets, including MSRA-10K [10], manually annotated Flickr natural image dataset, and so on.
- Each image in this saliency dataset is annotated with an accurate segmentation mask.
- In order to simulate the real-world inputs and learn a practical model, they drew holes on each image with arbitrary shapes randomly with a brush

# Contour Detection

- Uses DeepCut to detect the saliency objects in the image automatically.
- DeepCut is a CNN-based architecture that extracts and combines high-level and low-level features to predict a salient object mask with accurate boundaries.

# Contour Completion Module

- The goal of our contour completion module is to complete the missing contours of the input image that are inside the hole regions.
- The contour completion module is composed of a generator and a discriminator. The generator is a cascade of a coarse network and a refinement network.
- The coarse network is an encoder-decoder network with several convolutional and dilated convolutional layers.

# Contour Completion Module

- The predicted contours around the holes can be blurry and cannot be used as an effective guidance for the image completion module.
- To infer a more accurate contour, we adopt the refinement network which takes the coarse contour as input and output a cleaner and more precise contour.
- The refined contour is then fed to the contour discriminator for adversarial training.
- The contour discriminator is a fully convolutional PatchGAN discriminator that outputs a score map instead of a single score

# Image Completion Module

- Guided by the completed contours, the model gains the basic knowledge of where the foreground and background pixels are.
- The image completion module takes the incomplete image, the completed contour and the hole mask as inputs, and outputs the completed image.
- It shares a similar architecture as the contour completion module.
- The discriminator takes the generated image/ground truth image along with the hole mask indicating the location of the missing pixels as inputs and tells whether the input pair is real or fake.

# Experiments

Input PatchMatch Global&Local ContextAttention PartialConv GatedConv Ours Ground-Truth

# Experiments

Input PatchMatch Global&Local ContextAttention PartialConv GatedConv Ours Ground-Truth

# Quantitative Evaluation

- They randomly selected 500 images from the testing saliency dataset and generated both overlap and non-overlap holes for each image. Then they ran each method on the corrupted images to obtain the final results.

Table 1. Quantitative results on the saliency dataset.

| Method | L1 Loss | L2 Loss | PSNR | SSIM |
|---|---|---|---|---|
| PatchMatch [5] | 0.01386 | 0.004278 | 26.94 | 0.9249 |
| Global&Local [13] | 0.02450 | 0.004445 | 25.55 | 0.9005 |
| ContextAttention [29] | 0.02116 | 0.007417 | 24.01 | 0.9035 |
| PartialConv [17] | 0.01085 | 0.002437 | 29.24 | 0.9333 |
| GatedConv [28] | 0.009966 | 0.002531 | 29.26 | 0.9353 |
| Ours No Guided | 0.010002 | 0.002597 | 29.35 | 0.9356 |
| Ours Guided | **0.009327** | **0.002329** | **29.86** | **0.9383** |

# User study

- They randomly selected 50 images from the testing dataset, corrupted them with random holes and then obtained the inpainted results of each method.
- Then showed the results of each image to 22 users and asked them to select a single best result.

Table 2. User preference for the results of each method.

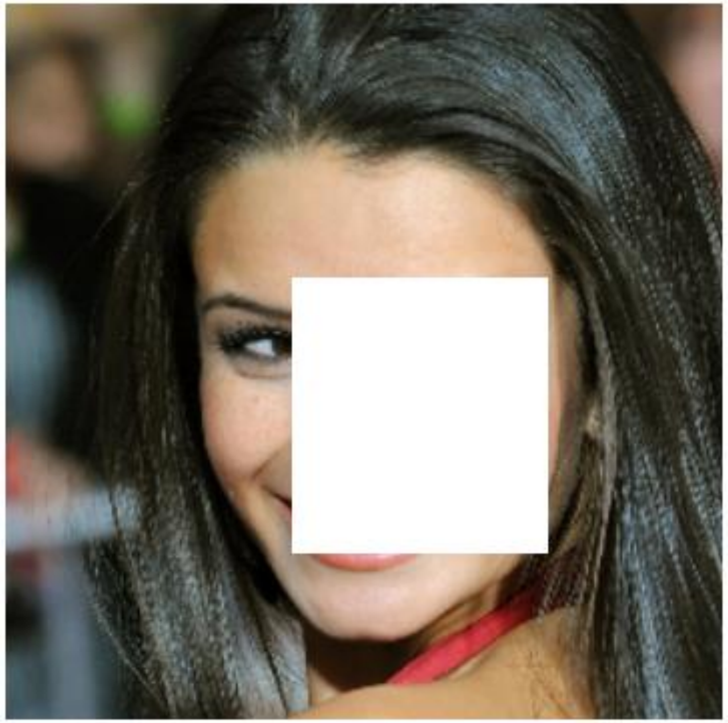
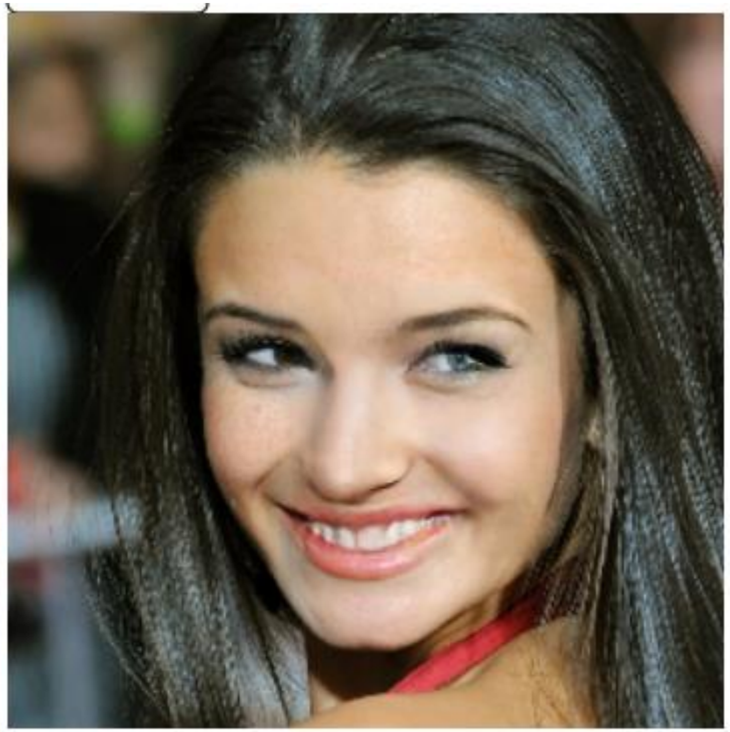
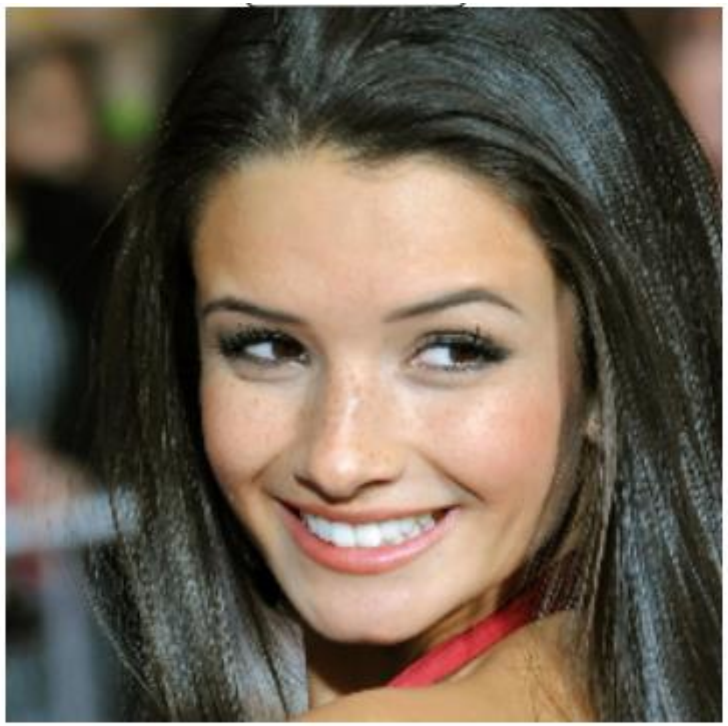
| Method | Preference Counts |
|---|---|
| PatchMatch [5] | 23 |
| Global&Local [13] | 5 |
| ContextAttention [29] | 4 |
| PartialConv [17] | 90 |
| GatedConv [28] | 100 |
| Ours No Guide | 146 |
| Ours Guided | **731** |

# Ablation Study

- Comparison of full model to the model without contour as guidance.

# Other examples

- Video
- Online demo (http://jiahuiyu.com/deepfill/)



# Conclusion

- In this paper has been presented the foreground-aware image inpainting model for challenging scenarios involving prediction of both foreground and background pixels.
- The model first detects and completes the contours of the foreground objects in the image, then uses the completed contours as a guidance to inpaint the image.
- The model significantly outperforms various state-of-the-art models both quantitatively and qualitatively.