Staging E-Commerce Products for Online Advertising using Retrieval Assisted Image Generation

Yueh-Ning Ku* (Snap),
Mikhail Kuznetsov* (Amazon),
Shaunak Mishra* (Amazon) – presenter,
Paloma de Juan (Yahoo Research)

*work done at Yahoo Research

AdKDD 2023
Advertising e-commerce products

e-commerce product catalog ➔ advertising platform ➔ publisher
Advertising e-commerce products

e-commerce product catalog → advertising platform → publisher

Top 5 home decor ideas

[organic content]

[ad content]
Advertising e-commerce products: staging

- e-commerce product catalog
- advertising platform
- publisher

Top 5 home decor ideas

- [organic content]
- ....
- [ad content]
- [organic content]

staged/lifestyle images blend well with page content, have higher engagement
Staging e-commerce products

stage physically?
◆ $$$
◆ can’t scale

can generative AI help?
Staging products using image generation

staged background generation

task 1

vanilla staging
(unstaged → staged via background gen.)
Staging products using image generation

vanilla staging
(unstaged → staged via background gen.)

copy-paste staging
(copy staging from other product images + in-paint)
Staging products using image generation

- **Staged background generation**

**Task 1**: Vanilla staging

- (unstaged → staged via background gen.)

**Task 2**: Copy-paste staging

- (copy staging from other product images + in-paint)

**Task 3 (bonus 😊)**: Image → parallax animation
Task 1: vanilla staging (using pix2pix)

Pix2pix [1] is a conditional generative adversarial network (GAN)

Task 1: vanilla staging (using pix2pix)

Pix2pix [1] is a conditional generative adversarial network (GAN)

Staging products using image generation: task 2

Staged background generation

Task 1

Vanilla staging
(unstaged → staged via background gen.)

Task 2

Copy-paste staging
(copy staging from other product images + in-paint)

Task 3 (bonus 😊)

Image → parallax animation
Copy-paste staging: core idea

product image with staging

using same staging for an unstaged product (red sofa)

in-painting focus on smaller regions easier
Copy-paste staging workflow (retrieval assisted gen.)

- pool of staged product images
- unstaged product image (input)

**Retrieve** similar product images with staging (inception-V3)

- remove products from similar images with staging

**In-paint** the empty mask in background after product removal

- copy-paste input product onto in-painted similar product background

sofa 1

sofa 2

sofa 1
Copy-paste staging workflow (retrieval assisted gen.)

- pool of staged product images

  - retrieve similar product images with staging (inception-V3)

  - remove products from similar images with staging

  - in-paint the empty mask in background after product removal

  - copy-paste input product onto in-painted similar product background

inspired from retrieval augmented generation (RAG) in text generation
Image inpainting

- Image inpainting is the task of reconstructing missing regions in an image, e.g. object removal, image restoration, manipulation.
- We propose to use an adapted EdgeConnect [2] model to fill the gap between the empty mask (from the staged product) and the target (unstaged) product.
  - EdgeConnect: generated edges and then generates color and texture

Image inpainting

- Image inpainting is the task of reconstructing missing regions in an image, e.g. object removal, image restoration, manipulation.
- We propose to use an adapted EdgeConnect [2] model to fill the gap between the empty mask (from the staged product) and the target (unstaged) product.
  - EdgeConnect: generated edges and then generates color and texture
  - **Our adaptation:** weighted boundary loss to focus on boundaries

---

Copy-paste staging demo
Copy-paste staging demo
Results: human evaluation

Human auditors were given three tasks (100 samples per task)

<table>
<thead>
<tr>
<th>audit task</th>
<th>score</th>
</tr>
</thead>
<tbody>
<tr>
<td>vanilla staging (pix2pix) better than ground truth</td>
<td>0%</td>
</tr>
<tr>
<td>copy-paste staging (our approach) better than ground truth</td>
<td>3%</td>
</tr>
<tr>
<td>copy-paste staging better than vanilla staging (pix2pix)</td>
<td>76%</td>
</tr>
</tbody>
</table>
Results: FID

Experiments on data from Yahoo (sample of ~ 2000 furniture product images).

Frechet Inception Distance (FID) [3] calculates the (feature distribution) distance between target domain and generated domain; **the smaller the better.**

<table>
<thead>
<tr>
<th>copy-paste staging</th>
<th>baseline FID (EdgeConnect)</th>
<th>our approach FID (EdgeConnect + weighted boundary loss)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>38.44</td>
<td><strong>37.44</strong></td>
</tr>
</tbody>
</table>

What about the bonus?

- **task 1**
  - vanilla staging
    (unstaged → staged via background gen.)

- **task 2**
  - copy-paste staging
    (copy staging from other product images + in-paint)

- **task 3 (bonus 😊)**
  - image → parallax animation
Image to parallax animation

Link to video:
https://www.dropbox.com/s/9at5gz24ukhf2gi/product_staging_image_to_parallax_demo.mp4?dl=0
Image to parallax animation

Link to video:
https://www.dropbox.com/s/9at5gz24ukhf2gi/product_staging_image_to_parallax_demo.mp4?dl=0
Conclusion

- staged background generation

Task 1
- vanilla staging
  (unstaged → staged via background gen.)

Task 2
- copy-paste staging
  (copy staging from other product images + in-paint)

Task 3 (bonus 😊)
- image → parallax animation

- copy-paste better than vanilla (FID, human eval.); need online test for further validation
- room for improvement in terms of shadows/lighting, hallucinations
- retrieval based ideas can be extended to recent stable diffusion based models