Which shoes fit this dress?
Using product images to infer “perfect pairings” across product categories without supervision

Luca Gaegauf
luca.gaegauf@bf.uzh.ch

Markus Meierer
markus.meierer@business.uzh.ch

University of Zurich[UZH]

Project summary & status

- Research question:

  Is it possible to infer product recommendations ***recommendations across product categories*** (i.e., recommend dress based on image of a shoe) by only leveraging implicit ***information of only product images*** (and no other information)?

- Approach:

  - Use ***generative adversarial networks*** trained in unsupervised fashion.

  - Assess the performance based on (1) comparison to various baselines and (2) surveying potential customers on the perceived quality of the recommendations.

- Status:

  Initial results are promising but we are not there yet.

Motivation

- Situation:

  Recommendation systems ***crucial*** for online business performance.

- Complication:
  - Existing recommendation system approaches suffer from a range of ***issues***:
    - Fast changing product range leads to cold start problem.
    - Expensive to collect, prepare, and maintain data.
    - Data privacy concerns.
  - Some online platforms experience these issues stronger than others (e.g., ebay.com or etsy.com).

- Possible solution:

  In contrast to structured product information, ***product images are always available*** and could possibly be leveraged to infer cross-category recommendations (without using any other information).

Related literature

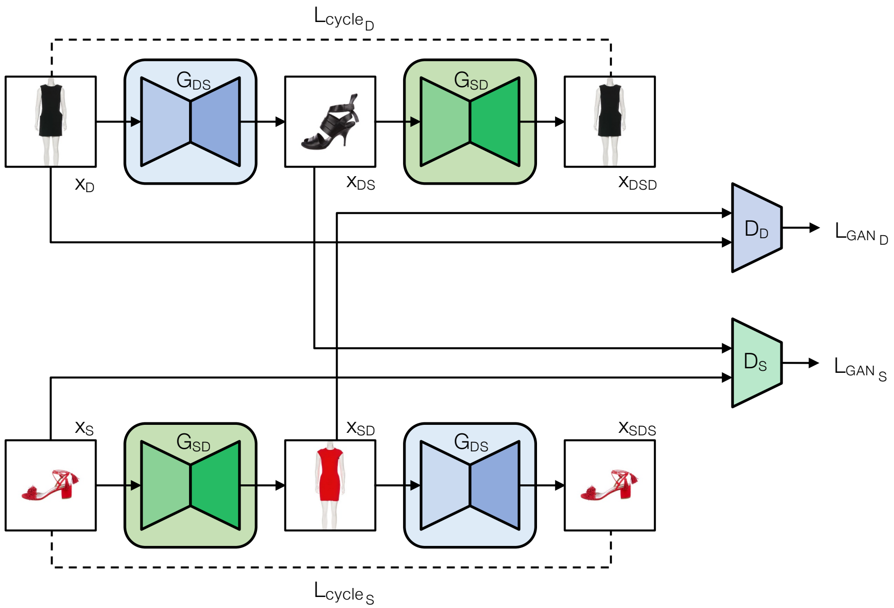
- Unsupervised image translation with generative adversarial networks (GANs):
  - DiscoGAN (Kim, et al.'s 2017)
  - CycleGAN (Zhu et al., 2017).
- Traditionally used techniques for recommender systems:
  - Collaborative filtering and content-based methods (see Breese et al, 2013).
  - AI-based recommendation systems (see Zhang et al., 2017):
    - Convolutional neural networks for within category recommendations (De Genova, 2017),
    - Next item recommendation with self attention (Zhang et al., 2018),
    - Deep content-based music recommendation (Van den Oord et al., 2013),
    - Liu et al. (2017) maps cross-domain feature spaces with similarity metric.

Kim, et al. (2017)

Approach: Step 1 – CycleGAN as an unsupervised image translation algorithm

Step 1: Apply CycleGAN (Zhu et al.'s 2017) :

- An unsupervised image translation algorithm,
- Uses unpaired data,
- Exhibits cycle-consistency.

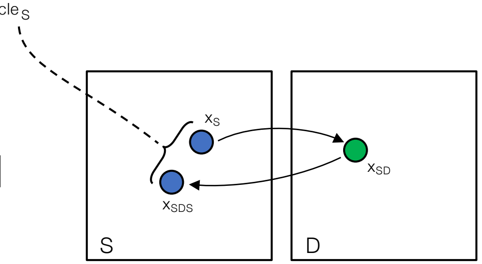


$$G^*_{DS}, G^*_{SD} = \arg \min_{G_{DS}, G_{SD}} \max_{D_D, D_S} L(G_{DS}, G_{SD}, D_D, D_S),$$

where $L(G_{DS}, G_{SD}, D_D, D_S) = L_{GAN_D} + L_{GAN_S} + \lambda(L_{cycle_D} + L_{cycle_S})$

$$L_{GAN_D} = L_{GAN}(G_{SD}, D_D, X_D, X_S) = \mathbb{E}_{x_D \sim p_{data}(x_D)} [\log D_D(x_D)] + \mathbb{E}_{x_S \sim p_{data}(x_S)} [\log (1 - D_D(G_{SD}(x_D)))]$$

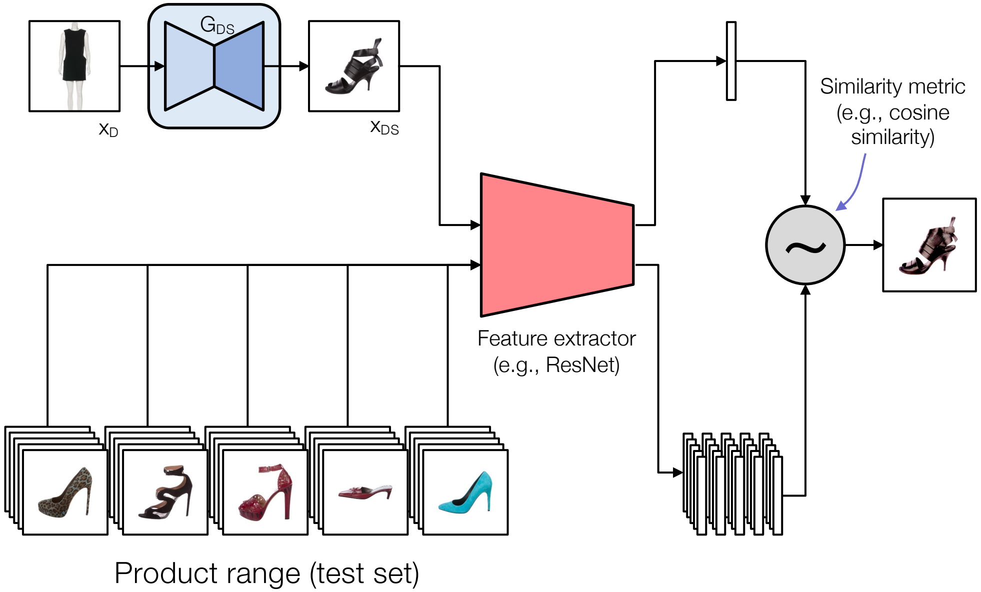
$$L_{cycle_D} = L_{cycle}(G_{SD}, G_{DS}, X_D) = \mathbb{E}_{x_D \sim p_{data}(x_D)} [\, || G_{DS}(G_{SD}(x_D)) - x_D ||_1 \,]$$



Approach: Step 2 – Using the trained generators to make recommendations

<u>Step 2: Generate recommendations:</u>

- Encode generated images using feature extractor.
- Encode images from product range using feature extractor.
- Rank product range according to similarity to generated images.
- Recommend products to customers based on ranking.



Data: Unpaired product images of shoes and dresses

Shoes (~100'000 images)

Input shoes:

Dresses (~100'000 images)

Input dresses:

→ Experiments also conducted on other fashion items (belts and handbags), accessories (earrings, bracelets, and rings), and furniture (chairs, couches, tables).

Images harvested from https://www.therealreal.com.

Results: Translating shoes to dresses and dresses to shoes

Results: Generating recommendations

Input image:  

Generated image:  

Results: Generating recommendations

Input image:

Generated image:

Top 3 recommendations:

Results: Generating recommendations

Input image:

Generated image:

Top 3 recommendations:

Top 3 benchmark recommendations:

Results: Generating recommendations

Input image:

Generated image:

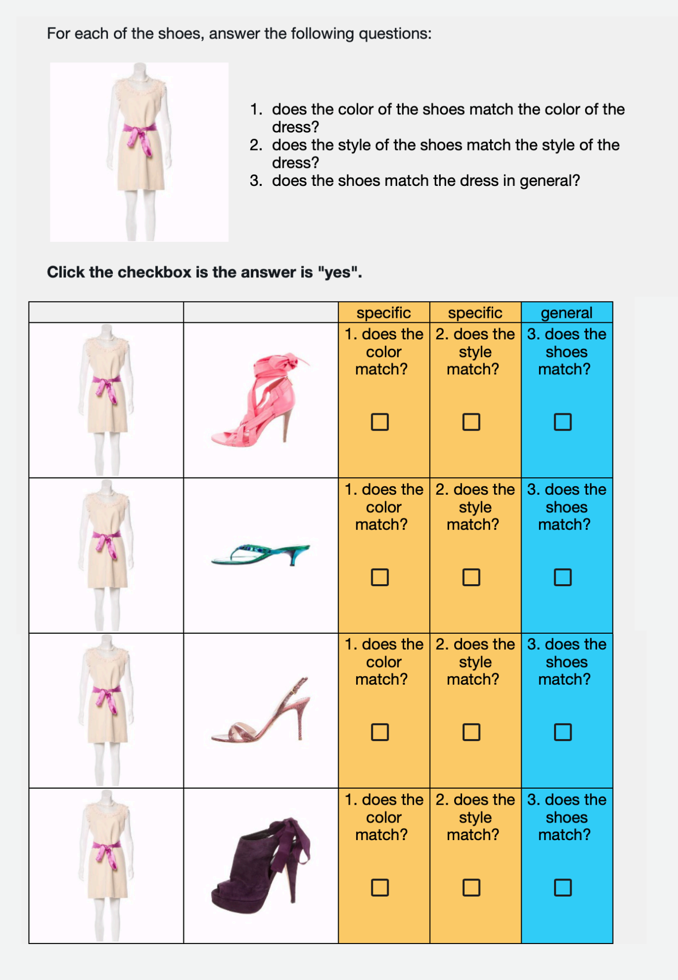
Top 3 recommendations:

Bottom 2 recommendations:

Top 3 benchmark recommendations:

Bottom 2 benchmark recommendations:

Evaluation: Surveying potential customers on Amazon Mechanical Turk



For each of the shoes, answer the following questions:

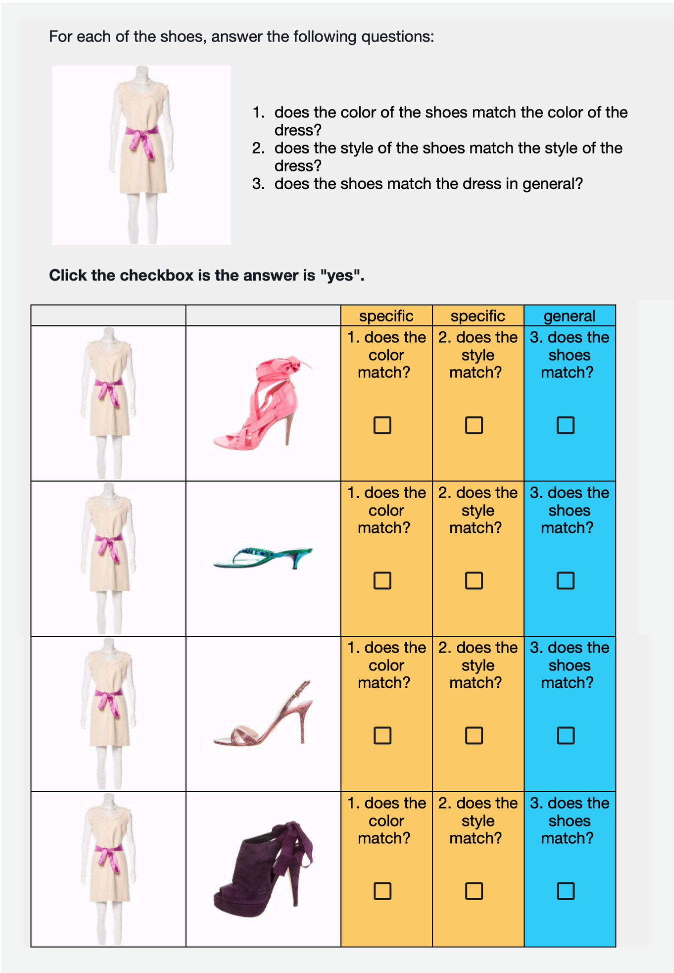
1. does the color of the shoes match the color of the dress?
2. does the style of the shoes match the style of the dress?
3. does the shoes match the dress in general?

Click the checkbox is the answer is "yes".

		specific	specific	general
		1. does the color match? ☐	2. does the style match? ☐	3. does the shoes match? ☐
		1. does the color match? ☐	2. does the style match? ☐	3. does the shoes match? ☐
		1. does the color match? ☐	2. does the style match? ☐	3. does the shoes match? ☐
		1. does the color match? ☐	2. does the style match? ☐	3. does the shoes match? ☐

		1. does the color match? ☐	2. does the style match? ☐	3. does the shoes match? ☐
		1. does the color match? ☐	2. does the style match? ☐	3. does the shoes match? ☐
		1. does the color match? ☐	2. does the style match? ☐	3. does the shoes match? ☐
		1. does the color match? ☐	2. does the style match? ☐	3. does the shoes match? ☐
		1. does the color match? ☐	2. does the style match? ☐	3. does the shoes match? ☐
		1. does the color match? ☐	2. does the style match? ☐	3. does the shoes match? ☐

Evaluation: Surveying potential customers on Amazon Mechanical Turk (preliminary results)



1. Does the color of the recommendation match the color of the source?

	GAN		Benchmark	
	top	bottom	top	bottom
Dress to shoes	59.5	26.8	47.4	25.5
Shoe to dresses	61.1	51.0	64.9	39.0

2. Does the style of the recommendation match the style of the source?

	GAN		Benchmark	
	top	bottom	top	bottom
Dress to shoes	72.0	20.2	45.1	40.3
Shoe to dresses	65.7	54.9	66.0	46.5

3. Does the recommendation match the source in general?

	GAN		Benchmark	
	top	bottom	top	bottom
Dress to shoes	58.6	16.5	32.4	23.7
Shoe to dresses	55.4	46.8	59.6	38.1

Promising results but we are not there yet

Summary:

- Unsupervised generative adversarial network based-recommendation system,
- Trained solely on unpaired product images available on retail websites,
- Recommendations display promising initial results, but we're not there yet.

Next steps:

- Pretraining individual components of the CycleGAN.
- Incorporate progressive growing (Karras et al., 2017) into the CycleGAN architecture.

References

- Breese, J. S., Heckerman, D., & Kadie, C. (2013). Empirical analysis of predictive algorithms for collaborative filtering. arXiv preprint arXiv:1301.7363.
- De Genova, V., (2017). Recommending visually similar products using content based features, https://tech.wayfair.com/data-science/2017/12/recommending-visually-similar-products-using-content-based-features/.
- Karras, T., Aila, T., Laine, S., & Lehtinen, J. (2017). Progressive growing of gans for improved quality, stability, and variation. arXiv preprint arXiv:1710.10196.
- Kim, T., Cha, M., Kim, H., Lee, J. K., & Kim, J. (2017). Learning to discover cross-domain relations with generative adversarial networks. arXiv preprint arXiv:1703.05192.
- Liu, Q., Wu, S., & Wang, L. (2017, August). DeepStyle: Learning user preferences for visual recommendation. In Proceedings of the 40th International ACM SIGIR Conference on Research and Development in Information Retrieval (pp. 841-844).
- Van den Oord, A., Dieleman, S., & Schrauwen, B. (2013). Deep content-based music recommendation. In Advances in neural information processing systems (pp. 2643-2651).
- Zhang, S., Tay, Y., Yao, L., & Sun, A. (2018). Next item recommendation with self-attention. arXiv preprint arXiv:1808.06414.
- Zhang, S., Yao, L., Sun, A., & Tay, Y. (2019). Deep learning based recommender system: A survey and new perspectives. ACM Computing Surveys (CSUR), 52(1), 1-38.
- Zhu, J. Y., Park, T., Isola, P., & Efros, A. A. (2017). Unpaired image-to-image translation using cycle-consistent adversarial networks. In Proceedings of the IEEE international conference on computer vision (pp. 2223-2232).