Visual Representation for Capturing Creator Theme in Brand-Creator Marketplace

Sarel Duanis
sarel@lightricks.com
Lightricks LTD
Jerusalem, Israel

Keren Gaiger
kgaiger@lightricks.com
Lightricks LTD
Jerusalem, Israel

Ravid Cohen
ravid@lightricks.com
Lightricks LTD
Jerusalem, Israel

Shaked Zychlinski
szychlinski@lightricks.com
Lightricks LTD
Jerusalem, Israel

Asnat Greenstein-Messica
asi@lightricks.com
Lightricks LTD
Jerusalem, Israel

ABSTRACT

Providing cold start recommendations in a brand-creator marketplace is challenging as brands' preferences extend beyond the mere objects depicted in the creator's content and encompass the creator's individual theme consistently thatresonates across images shared on her social media profile. Furthermore, brands often use textual keywords to describe their campaign's aesthetic appeal, with which creators must align. To address these challenges, we propose two methods: SAME (Same Account Media Embedding), a novel creator representation employing a Siamese network to capture the unique creator theme and OAAR (Object-Agnostic Adjective Representation), enabling filtering creators based on textual adjectives that relate to aesthetic qualities through zero-shot learning. These two methods utilize CLIP, a state-of-the-art language-image model, and improve it in addressing the aforementioned challenges.

CCS CONCEPTS

• **Information systems → Top-k retrieval in databases**; **Similarity measures**; **Personalization**; **Search interfaces**.

KEYWORDS

cold-start recommendations, two-sided marketplace

ACM Reference Format:
Sarel Duanis, Keren Gaiger, Ravid Cohen, Shaked Zychlinski, and Asnat Greenstein-Messica. 2023. Visual Representation for Capturing Creator Theme in Brand-Creator Marketplace. In *Seventeenth ACM Conference on Recommender Systems (RecSys '23), September 18–22, 2023, Singapore, Singapore.* ACM, New York, NY, USA, 4 pages. https://doi.org/10.1145/3604915.3610237

1 INTRODUCTION

Recently, social media has emerged as an essential marketing channel. Platforms like Popular Pays by Lightricks help brands improve their capability to identify content creators and influencers for joint campaigns on social media. As a brand-creator marketplace, Popular Pays offer brands to find creators in several ways: 'Lookalikes' - helping to find creators who resemble a given one, Search tool - enabling to find creators based on keywords or filters and Personal Recommendations where the marketplace vendor recommend on creators with a high likelihood for collaboration. Lightricks is a leading provider of video and image editing mobile apps with tens of millions of active creators. Following Lightricks' acquisition of Popular Pays, a significant opportunity has arisen to enhance the marketplace with millions of talented creators, while also providing a monetization opportunity for Lightricks' creators through collaboration with brands. To address this opportunity, an effective 'Lookalikes' service and personal recommendations of new (cold start) creators is required. The success of these algorithms relies on the features representing the recommended items, in this case, the creators, as cited in [7]. A visual representation that encapsulates the repeating theme and concept in the creator profile is crucial key in this scenario. Encouraged from the success of the CLIP model [5] in different "zero - shot" visual classification tasks, we decided to leverage the CLIP model for creators representation. However, as demonstrated in Figure 1, when using CLIP embeddings to measure similarity between creators, we found this representation effectively captures object similarity among creators. However, it fails to encapsulate the unique recurring theme of the creator. Additionally, when searching creators using a keyword by comparing it to the creators images within the CLIP space, as proposed by [5], we found out that the CLIP representation fails to encode aesthetic characteristics such as "pleasing", "intriguing" or "romantic" used to describe the brand or campaign aesthetic appeal. Figure 2 illustrates this challenge.

In this paper, we present how we overcome these challenges by developing **SAME (Same-Account-Media-Embedding)** - a novel creator representation based on a Siamese network, capturing the unique creator's theme streamlined across her social media posts and **OAAR (Object Agnostic Adjective Representation)** - a zero-shot learning of object-agnostic adjective representation to enable filtering candidate creators according to the campaign required aesthetic appeal reflected by the adjective.

RecSys '23, September 18–22, 2023, Singapore, Singapore

ACM ISBN 979-8-4007-0241-9/23/09.
https://doi.org/10.1145/3604915.3610237

2 APPROACH

2.1 SAME (Same Account Media Embedding)

SAME transforms the CLIP embeddings into a new embedding space where the similarity between embeddings corresponds to the likelihood of their origin from the same creator profile. When representing creators as an aggregation of their SAME image embeddings, we can identify other creators who are likely to post images relate to the same domain or theme. The architecture of SAME is a Multi-Layer Perceptron (MLP) with three hidden layers, that gets as input CLIP embedding and outputs a SAME embedding. We train the model as a Siamese [3] network using a triplet loss[8] where the positive examples in the dataset are images originated from the same profile and negative examples are images originated form two different profiles. The loss minimizes the dot product between the positives while maximizing it between the negatives. The training loss function is described by Equation 1

$$\mathcal{L} = \sum_{i,j \in U, j \neq i} max(d(a_i, p_i) - d(a_i, n_j) + margin, 0) \tag{1}$$

Where U is the set of all creators, a_i, p_i represent two distinct SAME output representations of images from creator **i**, n_j denotes the output SAME representation of an image from a different creator **j**, **d** is the dot product function and *margin* is a hyper parameter.

2.2 OAAR (Object Agnostic Adjective Representation)

To overcome the challenge of filtering candidates by their alignment with an aesthetic appeal of a certain brand or campaign, textually described with adjectives such as "pleasing" or "awesome", we developed the "Object Agnostic Adjective Representation" method. This method, manipulates the CLIP directional space to achieve a representation of adjective tokens that is independent from related objects information. We create this representation using generated text prompts, with no reliance on images, by observing that the difference between two CLIP embeddings of a text prompt which describes a photo with and without the adjective token reflects the adjective itself. For instance, subtracting the embeddings of the prompt "a photo of a dog" from those of the prompt "a vivid photo of a dog" isolates the "vividness" direction tied to the photo and not to the dog. We further detach the "vividness" direction from context by averaging it across all 100 class names in CIFAR-100 dataset [4] and repeat this for each desired adjective. Finally, we use a dot product to measure similarity between the adjective representation and the image embedding, indicating the image representation projection length onto the adjective representations.

3 EVALUATION

SAME Evaluation: Creator representations can be utilized in two previously discussed use cases: calculating creator similarity for the 'Lookalikes' service and serving as input for our 'cold start' creator recommendation algorithm designed to rank creators based on their relevance to a specific brand and campaign. To assess the effectiveness of the SAME creator representation for the 'Lookalikes' service, we designed a "creator similarity" task. This task ranks creator profiles based on their similarity to a specific query profile. Among tens of thousands of candidate creators, one is identical to the one provided in the query but is represented by a different subset of images. Figure 3a demonstrates the (Hit Rate)@K results of ranking the correct creator within the first k positions using SAME versus CLIP creator representations. As can be seen, the SAME creator representation significantly outperforms the CLIP creator representation. This implies that measuring similarity between creators in the SAME embedding space improves our capability to identify creators who are more likely to post similar content on their social media profiles. To assess the effectiveness of SAME representations in the context of 'cold start' creator recommendations, we trained two distinct models on a dataset comprising tens of thousands of collaborations between brands and creators facilitated through the platform. Each model was trained once using SAME creator representation as a content feature and again using CLIP representation. K-Nearest-Neighbors and CB2CF [1], aiming to map the creator content representation to its BPR [6] representation, were used as content-based recommendation models. Figure 3b and 3c show that the models trained on SAME representations outperform those trained on CLIP representations. Additionally, in this case, the KNN model outperforms the CB2CF model, assuming that the content representation of the creators possesses sufficient strength to accomplish the recommendation task. A qualitative evaluation of the SAME content representation compared to the CLIP content representation appears in Figure 1. For the qualitative evaluation presented in the paper, we used the "Influencer and Brand Dataset" [2] available for research purposes.[1]

OAAR Evaluation: In order to evaluate the effectiveness of OAAR in filtering by adjective keyword which describes the brand's aesthetic appeal, we present in figure 2 a visual comparison of top ranked creators by vanilla CLIP representation of the prompts "A pleasing photo" and "An intriguing photo," as opposed to the OAAR-derived representations for the "pleasing" and "intriguing" directions, respectively. Upon examining the images sourced from top-ranked creators using OAAR for "pleasing," it becomes apparent that these images are characterized by a harmonious color profile, eliciting a sense of tranquility and calm in the viewer. Referring to the three creators ranked highest for the "intriguing" direction, we can notice a collection of unusual images, inciting a sense of curiosity and interest.

4 PRESENTERS

Sarel Duanis, a lead data scientist and machine learning researcher at Lightricks, holds a BSc degree in Computer and Electrical Engineering from the Hebrew University. Contributes to the vision domain recommendations, his patented research on automated video editing forms the foundation of one of Lightricks apps.

Keren Gaiger, a data scientist in the recommendations at Lightricks, contributing to the development of recommendation engines for the various Lightricks applications. She recently completed her MSc degree at Tel Aviv University and her thesis work focused on the field of recommendation systems for sequential data.

[1] the paper was prepared for not-for-profit, scientific, research, educational, and scholarship purposes only for the Conference and not for the commercial sale or promotion of any product or service.

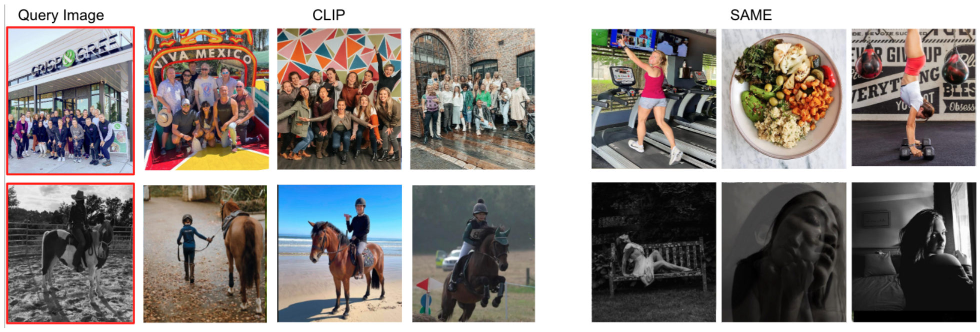


Figure 1: Top ranked images by similarity to query image sorted by vanilla CLIP and SAME representation, each image originated from a different profile. Notably, CLIP emphasizes similarity between objects, while SAME captures the theme or aesthetic style of the query image. In the first example, SAME identifies the theme of a healthy lifestyle and in the second, it discerns a black-and-white photographic style.

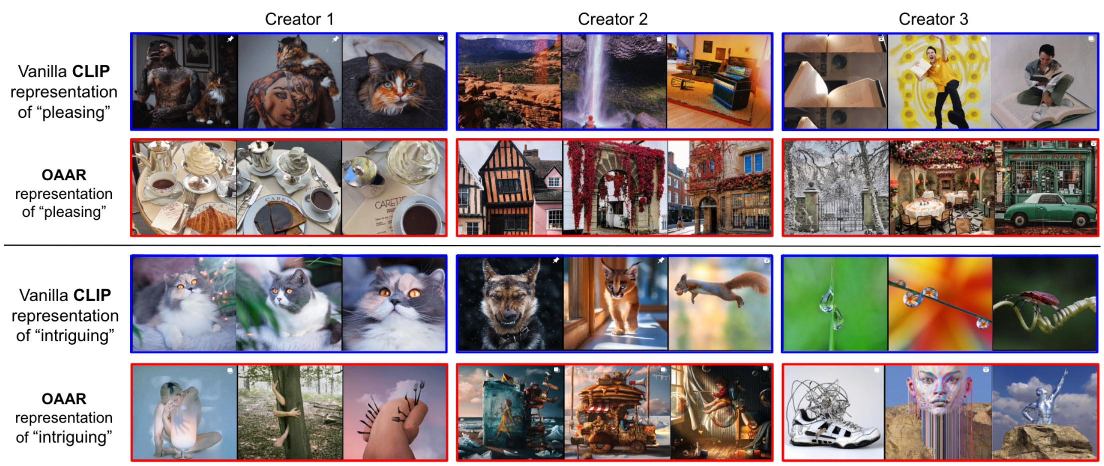
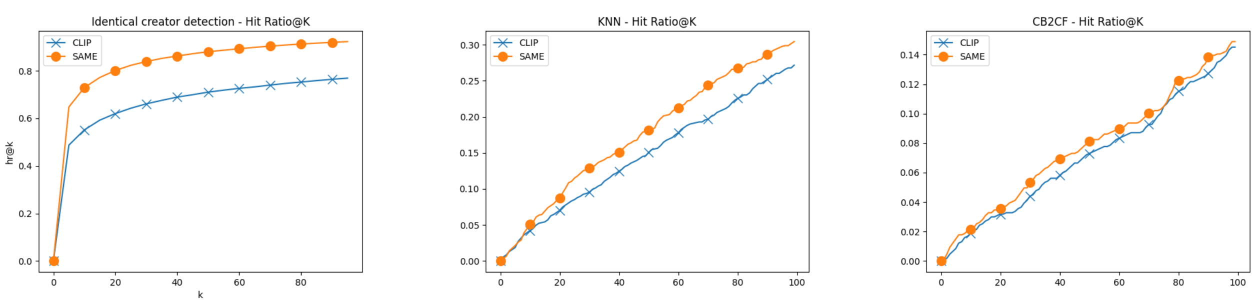


Figure 2: Top ranked creators identified as the most associated the Vanilla CLIP representations of "A pleasing photo" and "An intriguing photo" prompts comparing to the "pleasing" and "intriguing" OAAR representations, respectively.



(a) **(b)** **(c)**

Figure 3: Comparing between CLIP-based and SAME-based creator representations - 3a shows results of creators similarity task and 3b, 3c show results of cold start creator recommendations task using KNN and CB2CF models, respectively.

REFERENCES

[1] Oren Barkan, Noam Koenigstein, Eylon Yogev, and Ori Katz. 2019. CB2CF: a neural multiview content-to-collaborative filtering model for completely cold item recommendations. In *Proceedings of the 13th ACM Conference on Recommender Systems*. 228–236.

[2] Seungbae Kim, Jyun-Yu Jiang, and Wei Wang. 2021. Discovering undisclosed paid partnership on social media via aspect-attentive sponsored post learning. In *Proceedings of the 14th ACM International Conference on Web Search and Data Mining*. 319–327.

[3] Gregory Koch, Richard Zemel, Ruslan Salakhutdinov, et al. 2015. Siamese neural networks for one-shot image recognition. In *ICML deep learning workshop*, Vol. 2. Lille.

[4] Alex Krizhevsky, Geoffrey Hinton, et al. 2009. Learning multiple layers of features from tiny images. (2009).

[5] Alec Radford, Jong Wook Kim, Chris Hallacy, Aditya Ramesh, Gabriel Goh, Sandhini Agarwal, Girish Sastry, Amanda Askell, Pamela Mishkin, Jack Clark, et al. 2021. Learning transferable visual models from natural language supervision. In *International conference on machine learning*. PMLR, 8748–8763.

[6] Steffen Rendle, Christoph Freudenthaler, Zeno Gantner, and Lars Schmidt-Thieme. 2012. BPR: Bayesian personalized ranking from implicit feedback. *arXiv preprint arXiv:1205.2618* (2012).

[7] Francesco Ricci, Lior Rokach, and Bracha Shapira. 2010. Introduction to recommender systems handbook. In *Recommender systems handbook*. Springer, 1–35.

[8] Florian Schroff, Dmitry Kalenichenko, and James Philbin. 2015. Facenet: A unified embedding for face recognition and clustering. In *Proceedings of the IEEE conference on computer vision and pattern recognition (Lecture Notes in Computer Science)*. 815–823.