CausalRec: Causal Inference for Visual Debiasing in Visually-Aware Recommendation

Ruihong Qiu, Sen Wang, Zhi Chen, Hongzhi Yin, and Zi Huang
The University of Queensland
Brisbane, Australia
{r.qiu,sen.wang,zhi.chen,h.yin1}@uq.edu.au,huang@itee.uq.edu.au

ABSTRACT

Visually-aware recommendation on E-commerce platforms aims to leverage visual information of items to predict a user's preference for these items in addition to the historical user-item interaction records. It is commonly observed that user's attention to visual features does not always reflect the real preference. Although a user may click and view an item in light of a visual satisfaction of their expectations, a real purchase does not always occur due to the unsatisfaction of other essential features (e.g., brand, material, price). We refer to the reason for such a visually related interaction deviating from the real preference as a visual bias. Existing visually-aware models make use of the visual features as a separate collaborative signal similarly to other features to directly predict the user's preference without considering a potential bias, which gives rise to a visually biased recommendation. In this paper, we derive a causal graph to identify and analyze the visual bias of these existing methods. In this causal graph, the visual feature of an item acts as a *mediator*, which could introduce a spurious relationship between the user and the item. To eliminate this spurious relationship that misleads the prediction of the user's real preference, an *intervention* and a *counterfactual* inference are developed over the *mediator*. Particularly, the Total Indirect Effect is applied for a debiased prediction during the testing phase of the model. This causal inference framework is model agnostic such that it can be integrated into the existing methods. Furthermore, we propose a debiased visually-aware recommender system, denoted as CausalRec to effectively retain the supportive significance of the visual information and remove the visual bias. Extensive experiments are conducted on eight benchmark datasets, which shows the state-of-the-art performance of CausalRec and the efficacy of debiasing.

CCS CONCEPTS

• **Information systems** → **Recommender systems**.

KEYWORDS

visually-aware, causal inference, debiased recommendation

MM '21, October 20–24, 2021, Chengdu, China

ACM ISBN 978-1-4503-XXXX-X/18/06…$15.00
https://doi.org/10.1145/XXXX.XXXX

ACM Reference Format:
Ruihong Qiu, Sen Wang, Zhi Chen, Hongzhi Yin, and Zi Huang. 2021. CausalRec: Causal Inference for Visual Debiasing in Visually-Aware Recommendation. In *The 29th ACM International Conference on Multimedia (MM '21), October 20–24, 2021, Chengdu, China.* ACM, New York, NY, USA, 9 pages. https://doi.org/10.1145/XXXX.XXXX

1 INTRODUCTION

Visually-aware recommendation on E-commerce platforms takes the visual information of items into account in predicting a user's preference of these items in addition to the historical user-item interactions [14, 19, 28, 55]. Compared with traditional recommender systems, visually-aware methods improve the recommendation performance in many scenarios, e.g., shopping garments, where the users' preference is largely related to the appearance of items.

Although the visual feature is commonly used along with other features (e.g., brand, material, price) [14, 19, 28, 48, 55], the widely-used collaborative signal modeling of the visual feature is actually performing a biased learning scheme due to the visual feature itself. How could the visual feature give rise to bias while playing an important role in the recommendation? An example of the bias from the visual feature in buying white t-shirts is presented in Figure 1. Imagine that a user is looking for a white t-shirt made of cotton. When white t-shirts made of fabric or polyester are shown to the user, it is very likely for the user to click them because their appearance perfectly fits the user's need yet it will not lead to a purchase. These interaction records are imprecise for training the model for this user since these clicks do not reflect the real preference for the clicked items. Unfortunately, in most cases, all the clicks will be logged by the platform without discrimination. We refer to this mismatch between the interaction records and the real preference resulted by the visual feature as **visual bias**. Existing visually-aware recommender systems are mainly trained on visually biased records without debiasing procedure [14, 19, 28, 48, 55].

There exist various biases in recommendations, e.g., position bias, selection bias and popularity, which trigger the emergence of a few debiasing approaches correspondingly [1, 2, 10, 15, 22, 30, 32]. However, these approaches can hardly be applied to eliminate the visual bias, which originates from the item itself rather than the external bias mentioned above.

Recently, causal inference [33–36] has shown a great potential in removing the bias embedded in the data itself for vision-language tasks [37, 49, 51, 56]. Generally, in these methods, a causal graph is built to indicate a causal effect between different components for their tasks, where the causal effect quantifies the impact of a certain component on another one. To analyze the causal effect,

Figure 1: Example of visual bias on buying white t-shirts. A user is looking for a white t-shirt made of the suitable material. The user would click all of these white t-shirts of different materials since they look exactly like the target. However, the user's real preference relies on the material as well. With implicit feedback, it is difficult to tell if an interaction represents a real preference or just a spurious relationship purely between the visual feature and the interaction.

intervention and *counterfactual* inference are collectively common tools to provide debiased calculation results.

In light of the promising ability in removing bias of causal inference, we explore the way to adopt it for eliminating the visual bias in visually-aware recommendations. As a crucial step, we first identify the important factors in the recommendation: the user ID, the item ID, the visual feature of the item, the user-item preference match, the user's visual notice and the interaction. We introduce the user's notice of the visual feature of an item to indicate the user's pure visual preference without the influence by any other features. In many real-world shopping scenarios, the user's visual notice would strongly lead to user-item interactions when lacking other information (e.g., materials, brand, etc) for users' consideration at first glance of items. Therefore, it is expected to remove the causal effect of the visual notice in predicting the preference based on the biased interaction records. *Interventions* and the *counterfactual* inference are leveraged in this paper to pursue an unbiased prediction. The main idea is by asking the following question:

If a user had seen other items with the same visual feature, would this user still interact with these items?

The *counterfactual* thinking is shown by comparing the fact that the user has already interacted with an item and the imagination that the user "had seen other items with the same visual feature". After the comparison between these two situations, the direct visual effect is naturally identified since the visual feature is the only thing remaining unchanged. When this direct visual effect is eliminated, the prediction of the preference is expected to be visually debiased.

Specifically, in this paper, a causal graph is developed to analyze the visual bias in existing visually-aware recommendation methods. To perform the debiased recommendation for these methods, we propose to make use of the Total Indirect Effect (TIE) in the inference phase of these methods. Furthermore, a causal inference-based novel recommender model (CausalRec) is proposed to retain the supportive visual information and perform visual debiasing. The contributions of this paper are as follows:

- A causal inference-based framework is derived to identify, analyze and remove the visual bias in existing visually-aware recommender systems. To the best of our knowledge, this is the first attempt in this research area.
- A novel CausalRec model is proposed to unbiasedly make use of the visual feature whilst remove the visual bias in the visually-aware recommendation.
- Extensive experiments are conducted on eight datasets and the results demonstrate the efficacy of the causal inference module and the state-of-the-art performance of CausalRec.

2 PRELIMINARIES

In this section, the basic concepts of causal inference [33–36] are provided. In the following, capital letters are used for random variables and lowercase letters for an observation of random variables.

2.1 Causal Graph

A causal graph is a directed acyclic graph that represents the causal relationship between random variables. The causal graph is denoted as $\mathcal{G} = (\mathcal{V}, \mathcal{E})$, where $\mathcal{V}$ stands for a set of random variables (nodes) in the graph and $\mathcal{E}$ denotes the cause-and-effect relationships (edges) between those variables. Figure 2 (a) demonstrates an example of a causal graph, which includes three random variables A, B and C. In this figure, a few causal relationships can be identified. Since the variable A has a direct effect on another variable B, the causal path $A \rightarrow B$ indicates that A is a cause of B. Meanwhile, both A and B are the causes of C. If the causal effect of A towards C is to be investigated, there two corresponding causal paths linking A and C together: $A \rightarrow C$ and $A \rightarrow B \rightarrow C$, accounting for the direct effect and the indirect effect respectively.

When there is a treatment a for node A, it will has a causal effect on B and C so that they become B_a and C_{a,B_a} as in Figure 2 (a). If A is assigned to the value a^*, which stands for no-treatment in this paper with a null value or an average value for this variable [33–36], then B and C will become B_{a^*} and $C_{a^*,B_{a^*}}$ under this no-treatment. The shadowed nodes in Figure 2 (b) stand for the no-treatment.

2.2 Intervention

In causal inference, an *intervention* is an operation to cut off the incoming edges towards certain nodes. For example, in the causal graph in Figure 2 (a), if the direct effect of A on C is of interest to investigate, the causal path $A \rightarrow B \rightarrow C$ will become a spurious relationship since it introduces bias in the estimation of $P(C \mid A)$ according to Bayes rule: $P(C \mid A) = \sum_b P(C \mid A, b)\underline{P(b \mid A)}$, where we slightly abuse the notation $P(B = b) = P(b)$ and the *mediator* B introduces an observation bias through $P(b \mid A)$. If an *intervention* is exerted the node B to set a certain value b, i.e., $do(B = b)$ (simplified to $do(B)$), the causal path between A and B is cutoff. This omission of the edge is shown in Figure 2 (c) and (d). Since this *intervention* has eliminated the relationship between A and B, applying Bayes rule: $P(C \mid do(A)) = \sum_b P(C \mid A, b)\underline{P(b)}$. Here, B is not affected by A anymore, and vice versa, which requires to calculate the condition on b fairly. Note that Figure 2 (c) and Figure 2 (d) represent different meanings whether the treatment is on the direct or the indirect causal paths. For Figure 2 (c), the no-treatment $A = a^*$ is on the path $A \rightarrow C$, which results in C_{a^*,B_a}. While in Figure 2 (d), the treatment $A = a^*$ is on the path $A \rightarrow B \rightarrow C$, which results in $C_{a,B_{a^*}}$. The half shadowed node indicates the causes of it consist of both treatment and no-treatment.

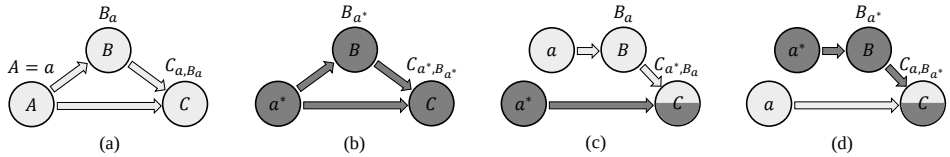


Figure 2: An example of intervention on causal graph. (a) The causal graph includes three random variables (nodes), A, B and C. The edges in the graph indicate the causal relationship between nodes. For example, the causal path $A \rightarrow B$ represents that A is the cause of B. And since there are two causal paths ($A \rightarrow C$ and $B \rightarrow C$) directing to C, both A and B are the causes of C. When there is an observation of $A = a$, B becomes B_a because it is based on A. Similarly, C becomes C_{a,B_a}. (b) A no-treatment assigns $A = a^*$ and this no-treatment results in C_{a^*,B^*}. (c) An intervention is operated on node B with a no-treatment to assign $A = a^*$ while leaving B unchanged. In this situation, the causal path of $A \rightarrow B$ is removed. The result of this no-treatment is $C_{a^*,B}$. (d) An intervention is operated on node B and a no-treatment $A = a^*$ affects the causal path $A \rightarrow B \rightarrow C$ instead of $A \rightarrow C$.

2.3 Counterfactual Notations

Counterfactual notations are used to translate the causal effect assumption from the causal graph to formulas. In the counterfactual situation, A is set to a different value a^* for different causal paths as in Section 2.2 above. Under this situation, C will become either C_{a^*,B_a} or $C_{a,B_{a^*}}$. These situations are called counterfactual because they do not really happen in the real world. It is imagined to investigate how a certain factor would affect the final outcome.

2.4 Causal Effect

Comparing the potential outcome after the counterfactual inference with the real outcome from an observation, the causal effect of the treatment used for the counterfact can be evaluated [16, 43]. Assume that Figure 2 (a) stands for under treatment $A = a$ and Figure 2 (b) stands for under no-treatment $A = a^*$. The Total Effect (TE) of the treatment $A = a$ on C is denoted as:

$$\text{TE} = C_{a,B_a} - C_{a^*,B_{a^*}}. \tag{1}$$

If the *intervention* is exerted according to Figure 2 (d), the Natural Direct Effect (NDE) can be derived as:

$$\text{NDE} = C_{a,B_{a^*}} - C_{a^*,B_{a^*}}. \tag{2}$$

Based on TE and NDE, total indirect effect (TIE) is defined as:

$$\text{TIE} = \text{TE} - \text{NDE} = C_{a,B_a} - C_{a,B_{a^*}}. \tag{3}$$

3 VISUAL BIAS IN VISUALLY-AWARE RECOMMENDATION

In this section, the causal view of existing models of visually-aware recommendation is discussed in detail.

3.1 Notation and Task Definition

In the following, lowercase letters are slightly overloaded to represent scalar, e.g., u for user ID and i for item ID. Bold lowercase letters are used to represent vectors, e.g., $\boldsymbol{\gamma}_u$ for latent embedding of user u and $\boldsymbol{\gamma}_i$ for latent embedding of item i. Bold capital letters are used to represent matrices and higher dimensional tensors, e.g., $\boldsymbol{V}_i$ for the image corresponding to item i. In the following causal graphs, the node set will include: item I, the corresponding visual feature V of the item, user U, the match M representing the real preference between the user and the item, the visual notice N of the user on the visual feature and the interaction Y.

The recommendation task considered in this paper only contains implicit feedback such as clicks and views instead of rating scores indicating the explicit preference. Within a recommendation scenario, there are a user set $\mathcal{U}$ and an item set $\mathcal{I}$. For each user u, the ID information is provided as well as the feedback to an item set $\mathcal{I}_u^+$. For each item i, besides the ID information, an image $\boldsymbol{V}_i$ of the item is also available to help with the prediction of user's preference. The objective of visually-aware recommendation is to generate a personalized item ranking for each user u over $\mathcal{I}$.

3.2 Non-visual Example: Matrix Factorization

Matrix Factorization (MF) has shown the state-of-the-art performance in recommendation tasks with the implicit feedback [21, 42]. The method is to develop a statistical model for the conditional probability $P(Y \mid I, U)$. A common usage of MF to predict the preference of a user u on an item i is formulated as follows:

$$y_{i,u} = \alpha + \beta_u + \beta_i + \boldsymbol{\gamma}_u^T \boldsymbol{\gamma}_i, \tag{4}$$

where α is an offset term, β_u and β_i are the bias terms of user and item respectively. $\boldsymbol{\gamma}_u$ and $\boldsymbol{\gamma}_i$ are the latent embedding factors of user u and item i respectively. The offset and bias terms are considered as mean effects of users and items. Latent embedding factors are performing a match between user's preference and item's properties in the form of dot product of dense vectors. The causal graph of MF is presented in Figure 3 (a), where it is clear that the user, the item and the match of the real preference are all the cause of an interaction with the causal paths: $U \rightarrow Y$, $I \rightarrow Y$ and $M \rightarrow Y$. Intuitively, the causal path $M \rightarrow Y$ is the wanted relationship to predict an interaction because it is based on the real preference.

Similar to the analysis in Section 2.1, both $I \rightarrow Y$ and $U \rightarrow Y$ are considered as backdoor paths of the causal path $M \rightarrow Y$. Therefore, I and U both are the *confounders*, which is also observed by Wei et al. [54]. Given that these two nodes have direct causal effect to the interaction, common situations account for them are the popularity bias of items and the active user bias.

3.3 Visual Bayesian Personalized Ranking

Visual Bayesian Personalized Ranking (VBPR) [14] is a strong baseline. The method targets at $P(Y \mid I, V, U)$ via extending the basic MF [42], which exploits the visual feature of the item similarly:

$$y_{i,v,u} = \alpha + \beta_u + \beta_i + \boldsymbol{\gamma}_u^T \boldsymbol{\gamma}_i + \boldsymbol{\theta}_u^T \left(\boldsymbol{E}\phi(\boldsymbol{V}_i)\right), \tag{5}$$

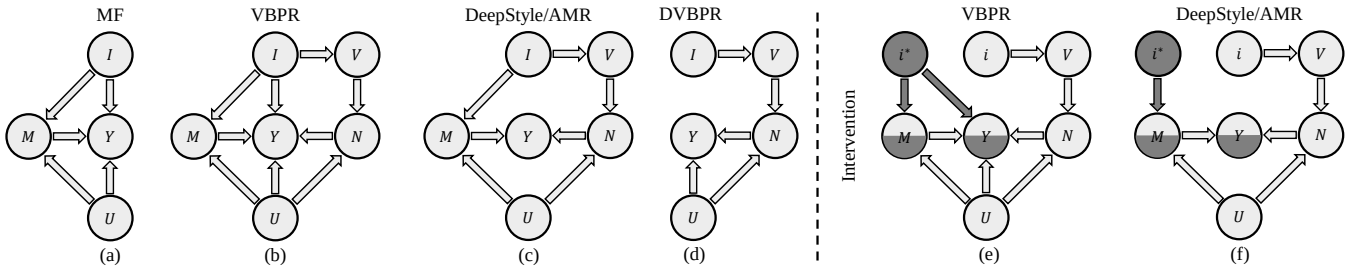


Figure 3: Variants of the causal graph of existing visually-aware recommender models and two examples of intervention. (a) MF leverages all of U, I and M to predict the interaction. (b) VBPR makes use of the direct effect of every component except the visual feature to predict the interaction. (c) DeepStyle and AMR remove the direct causal effect of U and I on the prediction of Y. (d) DVBPR further removes the direct effect of M, the match between the user and the item. Such a paradigm is applied more often on fashion compatibility. (e) The intervention of VBPR is conducted by setting a no-treatment for $I = i^*$. (f) The intervention of DeepStyle and AMR is conducted by setting a no-treatment for $I = i^*$.

where $\boldsymbol{E}$ is a transform matrix, ϕ is a backbone network (e.g., ResNet [12] and VGG [47]) to extract the visual feature representation from the item image $\boldsymbol{V}_i$ and $\boldsymbol{\theta}_u$ stands for a specific latent vector of the user towards the visual feature. The corresponding causal graph is shown in Figure 3 (b). Compared with the causal graph of MF, there are two extra nodes of the visual feature V and the visual notice N of the user towards the visual feature, where N can be thought of as $\boldsymbol{\theta}_u^T\left(\boldsymbol{E}\phi(\boldsymbol{V}_i)\right)$ in Equation (5). Furthermore, there is an extra direct cause of the interaction, $N \rightarrow Y$.

Within the causal graph, it can be concluded that the node V is the *mediator* and accountable for the visual bias. Such a visual bias is introduced by the pure visual notice as depicted as node N, which also lies on the backdoor path $M \leftarrow I \rightarrow V \rightarrow N \rightarrow Y$. VBPR also shares the same biases related to the item and the user itself as MF due to their direct causal effects towards the interaction.

3.4 DeepStyle and Adversarial Multimedia Recommendation

DeepStyle [28] and Adversarial Multimedia Recommendation [48] (AMR) are two follow-up methods of VBPR sharing the same causal graph with different techniques to improve the performance. They both remove the direct causal effects of I and U on Y.

The formulation of the prediction of DeepStyle is as follows:

$$y_{i,v,u} = \boldsymbol{\gamma}_u^T\left(\boldsymbol{E}\phi(\boldsymbol{V}_i) - \boldsymbol{c}_i + \boldsymbol{\gamma}_i\right), \tag{6}$$

where $\boldsymbol{c}_i$ represents the categorical information of the item image and subtracting this term from the visual feature is assumed to extract the more important style information. Furthermore, this method applies the same latent user vector $\boldsymbol{\gamma}_u$ to interact with both the visual feature and the item latent vector.

Similarly, AMR follows the same prediction paradigm while introducing a noise term to increase the robustness of the model:

$$y_{i,v,u} = \boldsymbol{\gamma}_u^T\left(\boldsymbol{E}\phi(\boldsymbol{V}_i) + \boldsymbol{\Delta}_i + \boldsymbol{\gamma}_i\right), \tag{7}$$

where $\boldsymbol{\Delta}_i$ denotes the noise added on the visual feature by as an adversary, which is trained in an adversarial learning style.

The causal graph of these two models is presented in Figure 3 (c). Compared with the causal graph of VBPR, it has the same set of nodes and removes two direct causal paths towards Y, $I \rightarrow Y$ and $U \rightarrow Y$. Different from VBPR, these two methods only have the backdoor path related to the visual notice, which indicates that DeepStyle and AMR are visually biased models rather than popularity biased and active user biased.

3.5 Deep Visual Bayesian Personalized Ranking

Deep Visual Bayesian Personalized Ranking (DVBPR) [19] is also based on VBPR. Although the visual feature is incorporated in DVBPR, the latent item vector is omitted, which is more related to outfit compatibility. Its prediction procedure is defined as:

$$y_{i,v,u} = \alpha + \beta_u + \boldsymbol{\theta}_u^T\left(\boldsymbol{E}\phi(\boldsymbol{V}_i)\right). \tag{8}$$

According to this equation, the only information related to the item is the visual feature. Therefore, in the causal graph of DVBPR in Figure 3 (d), there is no match node M. And the causes of the interaction consist of two causal paths, $U \rightarrow Y$ and $N \rightarrow Y$.

In terms of the outfit compatibility, the causal path $N \rightarrow Y$ is the base of prediction. Since U has a direct causal effect on Y as well, U is the *mediator* and it will introduce the active user bias.

4 VISUAL DEBIASING

In this section, the debiasing method based on causal inference and the CausalRec model are proposed. The main idea is to follow this question: *If a user had seen other items with the same visual feature, would this user still interact with these items?*

4.1 Counterfactual Inference in Visually-Aware Recommendation

In visually-aware recommendation, it is important to predict the match between user and item based on the real preference for the features of the item including the visual feature. The match is the criteria whether to recommend this item to the user. According to the previous analysis, there is visual bias resulted from a spurious relationship between the interaction and the user-item pair due to direct effect of the visual feature. Therefore, it is expected to remove this direct effect on the interaction. To further analyze the causes of the interaction, we describe the form of the interaction Y

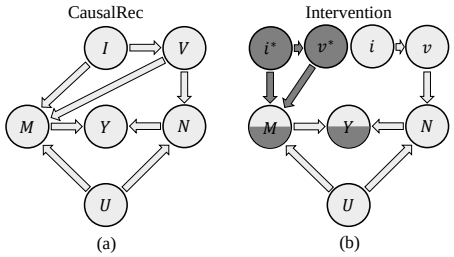


Figure 4: The causal graph of CausalRec and the corresponding intervention for visual debiasing. (a) The visual feature is used in both M and N. (b) In the intervention, both I and V are set to no-treatment for M.

based on a user u and an item i with the visual feature v as:

$$Y_{i,v,u} = Y(I = i, V = v, U = u) = Y_{M_{i,u},N_{v,u}}, \tag{9}$$

where M denotes the match and N stands for the visual notice in the causal graphs shown in Figure 3 (b) and (c). If a no-treatment $I = i^*$ is applied on both the direct and indirect effects, then the total effect (TE) of $I = i$ as defined in Equation (1) in Section 2.4 is:

$$\text{TE} = Y_{i,v,u} - Y_{i^*,v^*,u} = Y_{M_{i,u},N_{v,u}} - Y_{M_{i^*,u},N_{v^*,u}}. \tag{10}$$

4.1.1 Intervention. To eliminate the visual bias, it is expected to remove the direct effect of the visual feature on the interaction. In the causal graph, there are both direct and indirect effects of the visual feature. The direct effect lies on the causal path $I \rightarrow V \rightarrow N \rightarrow Y$. In the contrast, the visual feature can impact the match as well, which is the indirect effect in the causal path $I \rightarrow M \rightarrow Y$. According to our counterfactual thinking: "*If a user had seen other items with the same visual feature, would this user still interact with these items?*", the visual feature for the direct effect should be removed while stays in the indirect effect.

To investigate the direct effect of the visual feature, an *intervention* is conducted. The main purpose is to let the original visual feature affect the interaction with the direct effect while change the item for the indirect effect. Therefore, a no-treatment $I = i^*$ is exerted on the causal path of indirect effect as shown in Figure 3 (e) and (f). In this situation, the interaction is represented as:

$$Y_{i^*,v,u} = Y_{M_{i^*,u},N_{v,u}}. \tag{11}$$

Based on Equation (2), the natural direct effect of the visual feature of the treatment $I = i$ is:

$$\text{NDE} = Y_{i^*,v,u} - Y_{i^*,v^*,u} = Y_{M_{i^*,u},N_{v,u}} - Y_{M_{i^*,u},N_{v^*,u}}. \tag{12}$$

4.1.2 Counterfactual Inference. Since the *intervention* is already exerted on the causal graph, it is naturally to answer our counterfactual question by removing the direct effect of the visual feature on the interaction using TIE. According to Equation (3) in Section 2.4, the natural way is to use the Total Indirect Effect (TIE), i.e., minus the Natural Direct Effect (NDE) from the Total Effect (TE):

$$\text{TIE} = \text{TE} - \text{NDE} = Y_{i,v,u} - Y_{i^*,v,u} = Y_{M_{i,u},N_{v,u}} - Y_{M_{i^*,u},N_{v,u}}. \tag{13}$$

Using the maximum TIE for inference is different from the existing methods based on the conditional probability $P(Y \mid I, V, U)$.

Therefore, with Equation (5) of VBPR, the counterfactual prediction based on TIE becomes:

$$\hat{y}_{i,v,u} = \beta_i - \beta_{i^*} + \boldsymbol{\gamma}_u^T(\boldsymbol{\gamma}_i - \boldsymbol{\gamma}_{i^*}). \tag{14}$$

For DeepStyle and AMR, the counterfactual prediction becomes:

$$\hat{y}_{i,v,u} = \boldsymbol{\gamma}_u^T(\boldsymbol{\gamma}_i - \boldsymbol{\gamma}_{i^*}). \tag{15}$$

4.2 CausalRec Model

With the causal inference analysis for visual debiasing in the last section, after performing the debiasing procedure, the visual information will be fully removed. To effectively retain the supportive visual information and remove the visual bias, the CausalRec is proposed in this section.

4.2.1 Base Model. In light of the analysis of VBPR, DeepStyle and AMR, they can be unified as:

$$Y_{i,v,u} = \mathcal{F}(M_{i,u}, N_{v,u}), \tag{16}$$

where $\mathcal{F}$ represents a fusion function of the match M and the visual notice N, for example, a summation. Usually, both M and N are calculated with a dot product:

$$M_{i,u} = \boldsymbol{\gamma}_u^T \boldsymbol{\gamma}_i, \tag{17}$$

$$N_{v,u} = \boldsymbol{\theta}_u^T \boldsymbol{E}\phi(\boldsymbol{V}_i), \tag{18}$$

where $\boldsymbol{\theta}_u$ could be the same as $\boldsymbol{\gamma}_u$.

4.2.2 CausalRec Model. In addition to the base model, the visual indirect effect is included in the match node in the proposed CausalRec model as shown in Figure 4 (a). The detailed model is as follows:

$$M_{i,u} = \sigma(\boldsymbol{\gamma}_u^T \boldsymbol{\gamma}_i), \tag{19}$$

$$M_{i,v,u} = \sigma(\boldsymbol{\gamma}_u^T(\boldsymbol{\gamma}_i \circ \boldsymbol{E}\phi(\boldsymbol{V}_i)))), \tag{20}$$

$$N_{v,u} = \sigma(\boldsymbol{\theta}_u^T \boldsymbol{E}\phi(\boldsymbol{V}_i)), \tag{21}$$

$$\begin{aligned} Y_{i,v,u} &= \mathcal{F}(M_{i,u}, M_{i,v,u}, N_{v,u}) \\ &= M_{i,u} \cdot M_{i,v,u} \cdot N_{v,u}, \end{aligned} \tag{22}$$

where $\circ$ denotes the Hadamard product for the element-wise multiplication of vectors and σ denotes the Sigmoid function. In the choice of $\mathcal{F}$, a simple scalar multiplication is employed.

To train the model, we use the multitask learning framework to simultaneously train the CausalRec model with the following multi-tasking learning objective function:

$$\ell = \ell_{\text{rec}}(Y_{i,v,u}) + \ell_{\text{rec}}(N_{v,u}) + \ell_{\text{rec}}(M_{i,u}M_{i,v,u}), \tag{23}$$

where ℓ_{rec} is the BPR loss [42]:

$$\ell_{\text{rec}}(\hat{Y}) = \sum_{(u,i,j)\in O} -\ln \sigma\left(\hat{y}_{ui} - \hat{y}_{uj}\right) + \lambda_1 \|\Theta\|_2^2, \tag{24}$$

where $\hat{Y}$ is the prediction of interaction and O denotes the pairwise training dataset with i being the positive item and j being the negative item for user u. Θ represents all the trainable parameters and λ indicates the weight of this ℓ_2 regularization.

Table 1: Statistics of datasets

	♯ Users	♯ Items	♯ Interactions	Sparsity
Baby	19,822	7,776	163,856	99.89%
Beauty	25,837	16,893	227,920	99.95%
Clothing	58,197	44,310	422,474	99.98%
Grocery	16,318	11,581	165,893	99.91%
Office	6,913	4,775	68,306	99.79%
Sports	40,358	24,766	334,238	99.97%
Tools	20,134	14,774	163,451	99.95%
Toys	24,314	18,906	209,281	99.95%

4.2.3 CausalRec Inference. During the test phase of the CausalRec model, the counterfactual inference is applied with the *intervention* on the item. The *intervention* is detailed in Figure 4 (b). For CausalRec, the prediction can be elaborated as:

$$Y_{i,v,u} = Y_{M_{i,u},M_{i,v,u},N_{v,u}}. \tag{25}$$

Similarly, the no-treatment situation of $I = i^*$ is represented as:

$$Y_{i^*,v^*,u} = Y_{M_{i^*,u},M_{i^*,v^*,u},N_{v^*,u}}. \tag{26}$$

The final inference via TIE is as follows:

$$\text{TIE} = Y_{M_{i,u},M_{i,v,u},N_{v,u}} - Y_{M_{i^*,u},M_{i^*,v^*,u},N_{v,u}}. \tag{27}$$

To enhance the representation ability and retain a certain amount of the benevolent visual bias, a hyper-parameter λ_2 is used to control the scale of visual bias to be removed:

$$\hat{y}_{i,v,u} = M_{i,u} \cdot M_{i,v,u} \cdot N_{v,u} - \lambda_2 \cdot M_{i^*,u} \cdot M_{i^*,v^*,u} \cdot N_{v,u}. \tag{28}$$

5 EXPERIMENTS

In this section, extensive experiments will be conducted to evaluate the CausalRec model and the debiasing method. Four main research questions will be discussed:

- **RQ1**: Does CausalRec outperform the existing methods?
- **RQ2**: Does the causal inference-based debiasing method help with the existing methods?
- **RQ3**: How does different choices of implementation of the causal inference module help with removing the visual bias?
- **RQ4**: What is the sensitivity of hyper-parameters?

5.1 Experimental Setup

5.1.1 Datasets. The experiments are conducted on eight benchmark datasets: (1) Baby, (2) Beauty, (3) Clothing, Shoes and Jewelry (short for Clothing), (4) Grocery and Gourmet Food (short for Grocery), (5) Office Products (short for Office), (6) Sports and Outdoors (short for Sports), (7) Tools and Home Improvement (short for Tools) and (8) Toys on Amazon.com[1] [13, 29], which are widely used for visually-aware recommendation with available images for items [14, 19, 28, 29, 48]. The statistics of these datasets are presented in Table 1. The visual feature is extracted by a pretrained convolutional neural network following VBPR [14]. For all datasets, we consider the implicit feedback scenario[2]. Users and items that occur less than five times will be filtered out as well as the items without visual features.

[1] http://jmcauley.ucsd.edu/data/amazon/

[2] As one of the reviewers points out that it is relatively unfair to evaluate on a biased dataset. Yet, it is impractical to construct a debiased test set. Possible future solutions would include relying on the explicit ratings and A/B test.

Table 2: Overall performance.

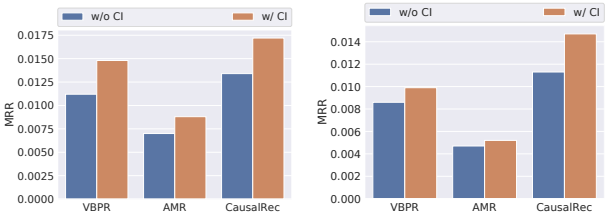
Dataset	Metric	BPR	VBPR	AMR	CausalRec	Improve
Baby	MRR	0.0146	0.0112	0.0070	**0.0172**	**17.81%**
	NDCG	0.0271	0.0215	0.0135	**0.0320**	**18.08%**
	HR	0.0983	0.0726	0.0462	**0.1047**	**6.51%**
Beauty	MRR	0.0071	0.0160	0.0102	**0.0192**	**20.00%**
	NDCG	0.0137	0.0315	0.0201	**0.0384**	**21.90%**
	HR	0.0476	0.1048	0.0690	**0.1254**	**19.66%**
Clothing	MRR	0.0038	0.0065	0.0036	**0.0088**	**35.38%**
	NDCG	0.0065	0.0125	0.0068	**0.0172**	**37.60%**
	HR	0.0216	0.0417	0.0235	**0.0528**	**26.62%**
Grocery	MRR	0.0140	0.0209	0.0147	**0.0250**	**19.61%**
	NDCG	0.0281	0.0414	0.0307	**0.0475**	**14.73%**
	HR	0.0981	0.1376	0.1068	**0.1541**	**11.99%**
Office	MRR	0.0152	0.0164	0.0145	**0.0223**	**35.98%**
	NDCG	0.0296	0.0340	0.0291	**0.0445**	**30.88%**
	HR	0.1040	0.1222	0.1021	**0.1537**	**25.78%**
Sports	MRR	0.0078	0.0086	0.0047	**0.0147**	**70.93%**
	NDCG	0.0140	0.0168	0.0087	**0.0258**	**53.57%**
	HR	0.0460	0.0563	0.0291	**0.0792**	**40.67%**
Tools	MRR	0.0110	0.0085	0.0048	**0.0134**	**21.82%**
	NDCG	0.0181	0.0162	0.0089	**0.0244**	**34.81%**
	HR	0.0537	0.0556	0.0324	**0.0775**	**39.39%**
Toys	MRR	0.0056	0.0106	0.0074	**0.0153**	**44.34%**
	NDCG	0.0106	0.0200	0.0141	**0.0305**	**52.50%**
	HR	0.0370	0.0663	0.0478	**0.1024**	**54.45%**

5.1.2 Metrics. To evaluate the performance of the recommender models, Mean Reciprocal Ranking (MRR), top-50 Normalized Discounted Cumulative Gain (NDCG) and top-50 Hit Ratio (HR) are used with a ranking of the whole item set for fair comparisons [23].

5.1.3 Implementation. There are a few hyper-parameters in the model. In the implementation of the model, we set the embedding size to 32 for the fairness of comparison. We use Adam [20] with a learning rate from {0.01,0.001,0.0001} and set the batch size as 100. λ_1 and λ_2 are chosen from {0.1,0.05,0.01,0.005} and {0,0.2,0.4,0.6,0.8,1,1.2} respectively. Our implementation is based on Cornac framework [45].

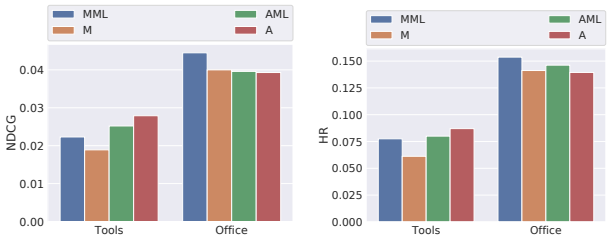
5.1.4 Baselines. The following baselines are used for comparisons and all of them have been discussed in detail in Section 3:

- **BPR** [42] is a baseline used only ID information of items and users. The visual feature is not included in this method.
- **VBPR** [14] is one of the earliest and the strongest baseline visually-aware recommender model. It extends the BPR method with a visual term multiplied with a user embedding to help with the collaborative signal.
- **AMR** [48] further simplifies the VBPR model by omitting the user and item bias terms. DeepStyle [28] shares a large similarity with AMR and thus be omitted here.



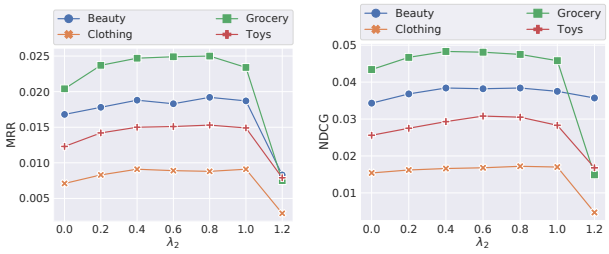
(a) Baby dataset. (b) Sports dataset.

Figure 5: Results of causal inference-based debiasing methods for visually-aware models.



(a) NDCG. (b) HR.

Figure 6: Results of different implementations of the causal inference module.



(a) MRR. (b) NDCG.

Figure 7: The parameter sensitivity of the λ_2.

5.2 RQ1: Overall Comparisons

The overall performance of baselines and the proposed CausalRec is presented in Table 2. In this experiment, we conduct the experiments on eight datasets and evaluate them with the following metrics: MRR, NDCG@50 and HR@50. According to the table, it is clear that the proposed CausalRec can achieve the best performance compared with all the baseline methods. The relative improvements are presented with the maximum one up to 70%.

As a collaborative filtering model using only ID information of users and items, BPR serves as a strong baseline in all datasets. Generally, VBPR can achieve a steady improvement compared with BPR. It is reasonable that the visual feature is important for these categories of items on the E-commerce platforms. Yet there are still some datasets, e.g., Baby and Tools, where the direct incorporation of the visual information will harm the recommendation performance. In addition to VBPR, AMR is a simplified version of VBPR. However, there is no significant improvement for AMR over VBPR.

Our proposed CausalRec model has higher performance for all datasets with the presented metrics. Compared with BPR, CausalRec explicitly includes a visual term to exploit the visual feature for recommendations, which is behaving similarly with the visual term in VBPR. While compared with VBPR and AMR, the most important difference lies in the causal inference module using the *counterfactual* thinking with the TIE quantification. With this module, the CausalRec model can perform a visual debiasing procedure while keeping the benevolent visual impact in the interaction.

5.3 RQ2: Visual Debiasing with Causal Inference

In this experiment, the proposed causal inference-based visual debiasing method is integrated into existing visually-aware recommendation models as well as the CausalRec model. This setting is designed to verify the generality of our debiasing method. VBPR and AMR are extended with the causal inference (CI) from Equation (14) and (15), denoted as VBPR w/ CI and AMR w/ CI. The biased version of CausalRec is shown here by not performing the CI in Equation (28), denoted as CausalRec w/o CI. The result is presented in Figure 5. The results are evaluated on Baby and Sports datasets with the MRR metric. Similar trends are observed on other datasets.

According to the figures, it is clear that our debiasing method can improve steadily for these existing methods as well as the CausalRec. This is mainly because the existing models are trained to simply maximize the interaction prediction probability. With our analysis in Section 3, their training methods are visually biased. Using a proper debiasing method, it is expected to eliminate the visual bias in the originally biased recommendation models.

5.4 RQ3: Different Causal Modules

In this experiment, we substitute the design of the multiplication-based fusing function $\mathcal{F}$ in Equation (22) with the original addition function used in CausalRec. The following variants named with the suffix A, M, AML and MML denote the addition fusing function, the addition fusing function with multi-tasking learning, the multiplication fusing function and the multiplication fusing function with multi-tasking learning respectively. The result is presented in Figure 6 for the CausalRec on Tools and Office datasets with NDCG and HR metrics. Similar phenomena are observed on other datasets.

In Figure 6, we can see that the choice of the fusing function will not affect too much of the performance. Either the proposed multiplication approach, M, or the addition based approach, A, are having comparable results. While the multitasking learning helps in a more general way for both the MML and AML. This could be due to the debiasing procedure is conducted by subtracting the causal effect of one of the branches in the model. If a branch is forced to learn more information about the recommendation task, subtracting this branch would give a more effective debiasing procedure.

5.5 RQ4: Parameter Sensitivity

In the CausalRec model, there are a few hyper-parameters as discussed in Section 5.1.3. Among all of them, the coefficient λ_2 in

Equation (28) is the one having direct control of the causal effect of the visual bias. The parameter sensitivity of λ_2 is tested in this section. We set the value of λ_2 in $\{0, 0.2, 0.4, 0.6, 0.8, 1, 1.2\}$. When $\lambda_2 = 0$, the model is not debiasing any visual feature. When $\lambda_2 = 1$, the direct causal effect of the visual feature is completely removed. The result is presented in Figure 7 with the evaluation on Beauty, Clothing, Grocery and Toys datasets over MRR and NDCG.

From the figures, it is clear that removing a certain scale of the direct causal effect of the visual feature can improve the recommendation result. As λ_2 increases, the recommendation performance will improve until completely removing the visual bias. If the visual bias is removed with a large scale factor, then the recommendation performance will be greatly harmed.

6 RELATED WORK

In this section, we review the related work of the visually-aware recommendation and causal inference.

6.1 Visually-Aware Recommendation

Visually-aware recommendations incorporates the visual feature of items into the prediction of the user's preference. Before the deep learning era, most methods depend on image retrieval for the recommendation [17, 18]. These methods assume that the user's preference for the similar visual feature would be similar. Kalantidis et al. [18] propose a method to firstly conduct a segmentation of the query image and retrieve items based on each of the predicted classes. In this work, the retrieval is conducted within the same class. In the following work, Jagadeesh et al. [17] find out that the semantic information of images is important and useful in the retrieval procedure. In their customized dataset setting, the semantic information is included with a large amount of annotations.

With more deep learning-based recommendation models being developed [38–41], recent methods can provide a more complicated modeling of the user-item interaction with the visual feature in addition to the simple retrieval-based approaches [7, 9, 14, 19, 28, 29, 31, 48, 57]. These methods mainly rely on pretrained deep learning framework to incorporate the visual knowledge, e.g., ResNet [12] and VGG [47]. IBR [29] recommends the complementary items based on the styles of item's visual feature. More generally, VBPR [14], AMR [48] and Fashion DNA [7] leverage the visual feature to support the collaborative filtering computation. With the help of the visual feature, these methods can improve the performance of the recommender systems in the sparse situation and the cold-start problem. DeepStyle [28] argues that the existing methods use the visual feature in an inappropriate way, in which the pretrained visual feature is majorly related to the class or category information. DeepStyle focuses more on the style of the visual feature rather than the categorical information. Besides passively using the existing visual features, DVBPR [19] applies an end-to-end trained CNN instead of the pretrained backbone for visual feature extraction. ImRec [31] proposes to use the reciprocal information between user groups with the aid of the image features.

There are a few existing methods focusing solely on the fashion recommendation task, which is more related to the compatibility of outfits [8, 11, 24–27, 50]. These models focus on recommending a suitable outfit and evaluating the compatibility of the outfit. The visually-aware recommendation has a broader scope than just the outfit. Many products on E-commerce platforms have important visual feature as well.

6.2 Causal Inference for Debiasing

In recent applications of machine learning to different tasks, causal inference has been used for debiasing towards different biases. In recommendation research area, the causal inference is mainly used to remove the interaction bias [3–6, 46], especially the popularity bias [53, 54]. The most widely used causal inference tool for these methods is Inverse Propensity Weighting (IPW) [44], which will conduct a reweighting on the interaction. A more recent work MACR [54] applies a causal graph [33–36] to analyze the causal effect of the popularity of items.

In multi-modal tasks, more and more methods make use of the causal inference to remove the bias in the data or in the model [37, 49, 51, 56]. For example, Tang et al. [49] use counterfactual inference in scene graph generation to remove the bias introduced by the image content. Qi et al. [37] argue that using *intervention* can remove the language bias in the historical language bias in the visual dialog. A more recent work investigates the clickbait issue via a causal graph method with the exposure feature being the source of the bias and the content feature is different from the exposure feature [52]. While in CausalRec, the visual feature serves as both the source of the visual bias and the content feature.

7 CONCLUSION

In this paper, the visual bias problem is identified and analyzed in the visually-aware recommendation. A novel causal inference framework is developed to investigate the direct and indirect causal effect of the visual feature of items on the interaction. To perform a debiased recommendation, the *intervention* and the *counterfactual* inference are applied after the biased training process. We further propose the CausalRec model to effectively make use of the visual feature and in the meanwhile to remove the visual bias. Extensive experiments are conducted on eight benchmark datasets, which demonstrates the state-of-the-art performance of the CausalRec model and the efficacy of the proposed visual debiased approach.

8 ACKNOWLEDGMENTS

The work was supported by Australian Research Council Discovery Project (ARC DP190102353, DP190101985, CE200100025)

REFERENCES

[1] Himan Abdollahpouri, Robin Burke, and Bamshad Mobasher. 2019. Managing Popularity Bias in Recommender Systems with Personalized Re-Ranking. In *FLAIRS*.

[2] Himan Abdollahpouri, Masoud Mansoury, Robin Burke, and Bamshad Mobasher. 2019. The Unfairness of Popularity Bias in Recommendation. In *RMSE@RecSys*.

[3] Aman Agarwal, Kenta Takatsu, Ivan Zaitsev, and Thorsten Joachims. 2019. A General Framework for Counterfactual Learning-to-Rank. In *SIGIR*.

[4] Alejandro Bellogín, Pablo Castells, and Iván Cantador. 2017. Statistical biases in Information Retrieval metrics for recommender systems. *Inf. Retr. J.* (2017).

[5] Stephen Bonner and Flavian Vasile. 2018. Causal embeddings for recommendation. In *RecSys*.

[6] Léon Bottou, Jonas Peters, Joaquin Quiñonero Candela, Denis Xavier Charles, Max Chickering, Elon Portugaly, Dipankar Ray, Patrice Y. Simard, and Ed Snelson. 2013. Counterfactual reasoning and learning systems: the example of computational advertising. *J. Mach. Learn. Res.* (2013).

[7] Christian Bracher, Sebastian Heinz, and Roland Vollgraf. 2016. Fashion DNA: Merging Content and Sales Data for Recommendation and Article Mapping. *CoRR* (2016).

[8] Zeyu Cui, Zekun Li, Shu Wu, Xiaoyu Zhang, and Liang Wang. 2019. Dressing as a Whole: Outfit Compatibility Learning Based on Node-wise Graph Neural Networks. In *WWW*.

[9] Xingzhong Du, Hongzhi Yin, Ling Chen, Yang Wang, Yi Yang, and Xiaofang Zhou. 2020. Personalized Video Recommendation Using Rich Contents from Videos. *IEEE Trans. Knowl. Data Eng.* (2020).

[10] Huifeng Guo, Jinkai Yu, Qing Liu, Ruiming Tang, and Yuzhou Zhang. 2019. PAL: a position-bias aware learning framework for CTR prediction in live recommender systems. In *RecSys*.

[11] Xintong Han, Zuxuan Wu, Yu-Gang Jiang, and Larry S. Davis. 2017. Learning Fashion Compatibility with Bidirectional LSTMs. In *ACMMM*.

[12] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. 2016. Deep Residual Learning for Image Recognition. In *CVPR*.

[13] Ruining He and Julian J. McAuley. 2016. Ups and Downs: Modeling the Visual Evolution of Fashion Trends with One-Class Collaborative Filtering. In *WWW*.

[14] Ruining He and Julian J. McAuley. 2016. VBPR: Visual Bayesian Personalized Ranking from Implicit Feedback. In *AAAI*.

[15] Katja Hofmann, Anne Schuth, Alejandro Bellogín, and Maarten de Rijke. 2014. Effects of Position Bias on Click-Based Recommender Evaluation. In *ECIR*.

[16] Guido Imbens and Donald Rubin. 1997. Bayesian Inference for Causal Effects in Randomized Experiments with Noncompliance. *Ann. Statist.* (1997).

[17] Vignesh Jagadeesh, Robinson Piramuthu, Anurag Bhardwaj, Wei Di, and Neel Sundaresan. 2014. Large scale visual recommendations from street fashion images. In *KDD*.

[18] Yannis Kalantidis, Lyndon Kennedy, and Li-Jia Li. 2013. Getting the look: clothing recognition and segmentation for automatic product suggestions in everyday photos. In *ICMR*.

[19] Wang-Cheng Kang, Chen Fang, Zhaowen Wang, and Julian J. McAuley. 2017. Visually-Aware Fashion Recommendation and Design with Generative Image Models. In *ICDM*.

[20] Diederik P. Kingma and Jimmy Ba. 2015. Adam: A Method for Stochastic Optimization. In *ICLR*.

[21] Yehuda Koren and Robert M. Bell. 2011. Advances in Collaborative Filtering. In *Recommender Systems Handbook*.

[22] Dominik Kowald, Markus Schedl, and Elisabeth Lex. 2020. The Unfairness of Popularity Bias in Music Recommendation: A Reproducibility Study. In *ECIR*.

[23] Walid Krichene and Steffen Rendle. [n.d.]. On Sampled Metrics for Item Recommendation. In *SIGKDD, year = 2020,*.

[24] Yuncheng Li, Liangliang Cao, Jiang Zhu, and Jiebo Luo. 2017. Mining Fashion Outfit Composition Using an End-to-End Deep Learning Approach on Set Data. *IEEE Trans. Multim.* (2017).

[25] Yang Li, Tong Chen, and Zi Huang. 2021. Attribute-aware Explainable Complementary Clothing Recommendation. *CoRR* (2021).

[26] Yang Li, Yadan Luo, and Zi Huang. 2020. Fashion Recommendation with Multi-relational Representation Learning. In *PAKDD*.

[27] Yusan Lin, Maryam Moosaei, and Hao Yang. 2020. OutfitNet: Fashion Outfit Recommendation with Attention-Based Multiple Instance Learning. In *WWW*.

[28] Qiang Liu, Shu Wu, and Liang Wang. 2017. DeepStyle: Learning User Preferences for Visual Recommendation. In *SIGIR*.

[29] Julian J. McAuley, Christopher Targett, Qinfeng Shi, and Anton van den Hengel. 2015. Image-Based Recommendations on Styles and Substitutes. In *SIGIR*.

[30] Marco Morik, Ashudeep Singh, Jessica Hong, and Thorsten Joachims. 2020. Controlling Fairness and Bias in Dynamic Learning-to-Rank. In *SIGIR*.

[31] James Neve and Ryan McConville. 2020. ImRec: Learning Reciprocal Preferences Using Images. In *RecSys*.

[32] Zohreh Ovaisi, Ragib Ahsan, Yifan Zhang, Kathryn Vasilaky, and Elena Zheleva. 2020. Correcting for Selection Bias in Learning-to-rank Systems. In *WWW*.

[33] Judea Pearl. 2001. Direct and Indirect Effects. In *UAI*.

[34] Judea Pearl. 2009. *Causality: Models, Reasoning and Inference*. Cambridge University Press.

[35] Judea Pearl, Madelyn Glymour, and Nicholas P Jewell. 2016. *Causal inference in statistics: A primer*. John Wiley & Sons.

[36] Judea Pearl and Dana Mackenzie. 2018. *The Book of Why: The New Science of Cause and Effect*. Basic Books, Inc.

[37] Jiaxin Qi, Yulei Niu, Jianqiang Huang, and Hanwang Zhang. 2020. Two Causal Principles for Improving Visual Dialog. In *CVPR*.

[38] Ruihong Qiu, Zi Huang, Tong Chen, and Hongzhi Yin. 2021. Exploiting Positional Information for Session-based Recommendation. *CoRR* (2021).

[39] Ruihong Qiu, Zi Huang, Jingjing Li, and Hongzhi Yin. 2020. Exploiting Cross-session Information for Session-based Recommendation with Graph Neural Networks. *ACM Trans. Inf. Syst.* (2020).

[40] Ruihong Qiu, Jingjing Li, Zi Huang, and Hongzhi Yin. 2019. Rethinking the Item Order in Session-based Recommendation with Graph Neural Networks. In *CIKM*.

[41] Ruihong Qiu, Hongzhi Yin, Zi Huang, and Tong Chen. 2020. GAG: Global Attributed Graph Neural Network for Streaming Session-based Recommendation. In *SIGIR*.

[42] Steffen Rendle, Christoph Freudenthaler, Zeno Gantner, and Lars Schmidt-Thieme. 2009. BPR: Bayesian Personalized Ranking from Implicit Feedback. In *UAI*.

[43] James Robins. 1986. A new approach to causal inference in mortality studies with a sustained exposure period—application to control of the healthy worker survivor effect. *Mathematical Modelling* (1986).

[44] Paul Rosenbaum and Donald Rubin. 1983. The central role of the propensity score in observational studies for causal effects. *Biometrika* (1983).

[45] Aghiles Salah, Quoc-Tuan Truong, and Hady W Lauw. 2020. Cornac: A Comparative Framework for Multimodal Recommender Systems. *Journal of Machine Learning Research* (2020).

[46] Tobias Schnabel, Adith Swaminathan, Ashudeep Singh, Navin Chandak, and Thorsten Joachims. 2016. Recommendations as Treatments: Debiasing Learning and Evaluation. In *ICML*.

[47] Karen Simonyan and Andrew Zisserman. 2015. Very Deep Convolutional Networks for Large-Scale Image Recognition. In *ICLR*.

[48] Jinhui Tang, Xiaoyu Du, Xiangnan He, Fajie Yuan, Qi Tian, and Tat-Seng Chua. 2020. Adversarial Training Towards Robust Multimedia Recommender System. *IEEE Trans. Knowl. Data Eng.* (2020).

[49] Kaihua Tang, Yulei Niu, Jianqiang Huang, Jiaxin Shi, and Hanwang Zhang. 2020. Unbiased Scene Graph Generation From Biased Training. In *CVPR*.

[50] Andreas Veit, Balazs Kovacs, Sean Bell, Julian J. McAuley, Kavita Bala, and Serge J. Belongie. 2015. Learning Visual Clothing Style with Heterogeneous Dyadic Co-Occurrences. In *ICCV*.

[51] Tan Wang, Jianqiang Huang, Hanwang Zhang, and Qianru Sun. 2020. Visual Commonsense R-CNN. In *CVPR*.

[52] Wenjie Wang, Fuli Feng, Xiangnan He, Hanwang Zhang, and Tat-Seng Chua. 2021. Clicks can be Cheating: Counterfactual Recommendation for Mitigating Clickbait Issue. In *SIGIR*.

[53] Yixin Wang, Dawen Liang, Laurent Charlin, and David M. Blei. 2020. Causal Inference for Recommender Systems. In *RecSys*.

[54] Tianxin Wei, Fuli Feng, Jiawei Chen, Chufeng Shi, Ziwei Wu, Jinfeng Yi, and Xiangnan He. 2021. Model-Agnostic Counterfactual Reasoning for Eliminating Popularity Bias in Recommender System. In *SIGKDD*.

[55] Bin Wu, Xiangnan He, Yun Chen, Liqiang Nie, Kai Zheng, and Yangdong Ye. 2020. Modeling Product's Visual and Functional Characteristics for Recommender Systems. *IEEE Transactions on Knowledge and Data Engineering* (2020).

[56] Zhongqi Yue, Tan Wang, Hanwang Zhang, Qianru Sun, and Xian-Sheng Hua. 2021. Counterfactual Zero-Shot and Open-Set Visual Recognition. In *CVPR*.

[57] Yan Zhang, Hongzhi Yin, Zi Huang, Xingzhong Du, Guowu Yang, and Defu Lian. 2018. Discrete Deep Learning for Fast Content-Aware Recommendation. In *WSDM*.