

基于视觉分割和语义相似性的趋势感知时尚推荐

Mohamed DJILANI
University of Luxembourg
Luxembourg
mohamed.djilani@uni.lu

Nassim ALI OUSALAH
University of Luxembourg
Luxembourg
nassim.alioussalah@uni.lu

Nidhal Eddine CHENNI
University of Luxembourg
Luxembourg
nidhal.chenni@uni.lu

Abstract

我们介绍了一种趋势感知和视觉基础的时尚推荐系统，该系统集成了深度视觉表示、服装感知分割、语义类别相似性和用户行为模拟。我们的流程通过语义分割对非服装区域进行遮罩来提取集中的视觉嵌入，随后使用预训练的 CNN 主干 (ResNet-50、DenseNet-121、VGG16) 进行特征提取。为了模拟现实的购物行为，我们生成受用户特定趋势性和物品流行度影响的合成购买历史。推荐是通过一个加权评分函数计算得出的，该函数融合了视觉相似性、语义一致性和流行度对齐。在 DeepFashion 数据集上的实验表明性别对齐一致且改善了类别相关性，其中 ResNet-50 实现了 64.95 % 类别相似性和最低流行度 MAE。消融研究确认了视觉和流行度提示的互补作用。我们的方法提供了一个可扩展的个性化时尚推荐框架，平衡个人风格和新兴趋势。我们的实现可以在 <https://github.com/meddjilani/FashionRecommender> 找到。

Keywords

Fashion Recommendation, Visual Similarity, Deep Learning, Content-Based Filtering

1 介绍

电子商务平台和在线零售的激增显著改变了时尚行业，为消费者提供了前所未有的选择。然而，在这些庞大的选择中导航以找到合适的服装仍然具有挑战性，这促使了个性化推荐系统的兴起。时尚推荐系统利用先进技术，特别是机器学习和深度学习，来建议符合个体偏好、风格和背景需求的产品。

卷积神经网络 (CNNs) [8] 的最新进展，通过高效捕获服装图像中固有的丰富视觉属性，使时尚推荐系统得到了显著的改进。利用诸如 ResNet50 [5]、DenseNet121 [4] 和 VGG16 [11] 等预训练模型，这些系统提取视觉嵌入，使基于视觉相似性的有意义和准确的推荐成为可能。除了视觉特征之外，诸如产品类别、性别和流行度等元数据以及时尚项目之间的语义关系，在改进推荐准确性方面也起着关键作用。

本文提出了一种综合的时尚推荐框架，该框架将预训练的 CNN 用于视觉特征提取，与多因素内容推荐引擎相结合。所提方法通过生成合成数据模拟真实的零售环境和用户行为，从而增强评价的稳健性。我们的方法通过结合视觉相似性、商品受欢迎程度、语义类别关系和用户特定属性（如性别偏好和流行趋势）来解决传统推荐方法的局限性。

本文的剩余部分组织如下：第二节讨论了时尚推荐系统领域的相关文献，重点介绍了最近的进展和尚存的挑战。第三节概述了所提出的方法，包括视觉特征提取、元数据模拟、合成用户购买生成、语义相似性度量和推荐评分的详细描述。第四节描述了系统的实施和评估，第五节对研究结果和未来的研究方向进行了总结。

2 相关工作

推荐系统在各个行业中变得越来越重要，这是因为它们通过提供个性化建议以迎合个人偏好来提升用户体验的能力。在其核心，这些系统分析用户交互的模式，以推荐可能对用户感兴趣的项目。最初，协同过滤作为一种流行的技术出现，依靠用户偏好之间的相似性生成建议，主要依赖于用户与项目交互的历史数据。然而，协同过滤在处理稀疏数据的问题时表现出困难——在这种情况下，许多项目很少有交互——以及冷启动问题，即由于数据不足，对新用户或新项目的推荐变得具有挑战性 [2]。

为了克服这些限制，开发了基于内容的方法。这些方法利用了项目特定的属性，如描述、类别和视觉特征以及用户的个人信息，来提供更相关的推荐。与协同过滤不同，基于内容的方法可以有效地管理新项目，因为推荐直接依赖于项目的特征，而不仅仅是用户的互动。

最近，将多模态表示学习整合到一起——结合多种类型的数据，如文本和图像——在推荐系统中显示出可喜的结果。例如，Yilma 和 Leiva [14] 表明结合项目的文本描述和视觉图像可以通过更好地捕捉单一模态可能忽略的复杂关系，显著提升推荐质量。在进一步扩展该研究时，Yilma 等人 [15] 研究了整合这些多模态特征的方法，特别指出早期融合策略——即在模型的早期阶段将来自多模态的数据组合——往往比后期融合策略表现更好，从而强调联合多模态数据表示的优势。

多利益相关者的视角也成为了现代推荐系统研究中的一个关键因素。通常，推荐系统主要关注终端用户；然而，Yilma 等人提出的 MOSAIC 框架也将内容提供商、策展人和平台运营商等其他利益相关者的目标和偏好纳入考虑。此方法确保推荐不仅令用户满意，同时也对其他相关方有益，可能提高推荐系统的实际可用性和整体认可度。

此外，考虑上下文——例如位置、时间和特定的环境限制——由于在提高推荐准确性和相关性方面的作用而受到关注。Yilma 等人引入了网络物理社会系统 (CPSS) 的概念，强调了在推荐过程中考虑现实世界上下文和限制的重要性。这有助于确保推荐不仅在理论上相关，而且在特定环境中切实适用和可执行。

此外，推荐系统在超越典型商业应用方面表现出了潜力，尤其是在治疗环境中。例如，Yilma 等人 [13] 探索了使用个性化的视觉艺术推荐作为治疗手段来提高心理健康状况，说明个性化推荐如何显著影响情感和心理健康的结果。

特别是在时尚领域，推荐系统经历了相当大的创新。Patil 等人 [1] 采用了 ResNet-50，这是一种以能够捕捉复杂视觉模式而闻名的深度学习架构，来开发一种推荐系统，能够识别时尚物品的重要视觉属性，如颜色、纹理和风格。类似地，Sivaranjani 等人 [12] 利用基于卷积神经网络 (CNN) 的视觉嵌入结合 K 近邻 (KNN) 算法来提高检索精度，超越传统的推荐方法。深度学习架构，特别是 ResNet，有效地解决了常见的神经网络问题，例如梯度消失，使其在基于图像的时尚推荐任务中极为有效。

多模态数据的整合进一步优化了时尚推荐系统。Varma 等人 [7] 介绍了一个 AI 驱动的系统, StyleSync, 该系统通过深度学习对服装类型和颜色进行分类, 并提供针对特定场合的个性化服装推荐。他们的方法强调了用户友好的界面, 并通过用户反馈不断进行改进, 显著提升了用户满意度和系统适应能力。

尽管有这些进步, 时尚推荐系统仍然面临挑战, 尤其是在平衡高度个性化的推荐与更广泛的市场趋势和库存可用性、地理位置等现实约束之间。我们提出的系统试图通过集成复杂的视觉特征提取技术、现实的用户模拟、语义相似性评估和地理考虑, 来解决这些问题, 旨在优化个性化和实际应用于真实世界场景的可行性。

3 提出的方法

我们提出了一种基于内容的时尚推荐系统, 该系统利用深度视觉特征、类别语义和动态行为模拟来生成个性化的产品建议。该架构结合了服装级别的分割、多个 CNN 骨干网以及一个现实的合成购买模拟器, 以模拟用户在不同时尚趋势下的行为。系统概述如图 1 所示。

3.1 服装感知的视觉特征提取

为了专注于衣物项目, 我们使用语义分割掩码预处理图像以去除非服装区域 (例如, 脸部、头发、皮肤)。我们在分割结果上应用基于颜色的掩码, 并使用预训练的 CNN 骨干网络从掩码后的图像中提取视觉特征。具体而言, 我们采用了 ResNet50 [3], DenseNet121 [4], 以及 VGG16 [11]。每个模型输出一个一维特征向量, 代表产品图像的视觉外观。我们的服装感知掩码策略的一个示例如图 2 所示。

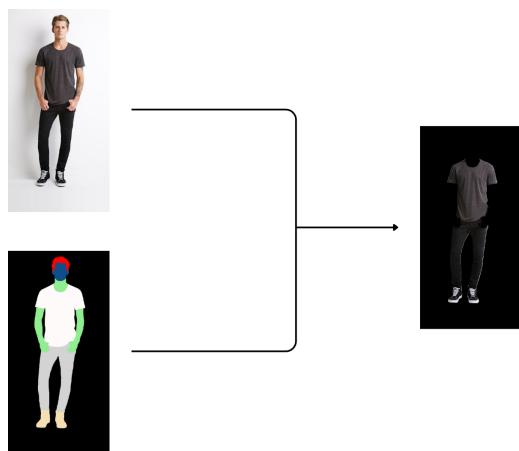


Figure 2: 我们提出的服装感知遮罩策略概述, 旨在更好地聚焦于仅与服装相关的特征提取。

3.2 商品元数据与商店分配

每个产品图像在其文件名中编码了元数据, 包括性别、类别和一个唯一的物品 ID。使用正则表达式来解析和构造这些信息。为了模拟全渠道零售体验, 物品被分配到均匀分布在地理空间中的合成店铺位置。每个店铺被分配了随机的坐标 (纬度, 经度), 这些坐标后来用于靠近感知推荐系统。

3.3 趋势感知合成购买生成

为了在真实的行为模式下评估推荐质量, 我们模拟了 M 用户一整年的购买历史。每个用户被分配一个趋势得分 $t \in [0, 1]$, 以反映他们对流行物品的喜好。物品的人气最初是随机初始化的, 并在购买发生时动态更新。

购买历史生成遵循以下步骤:

- 一个随机项目引发用户的性别偏好。
- 对于后续购买, 商品的采样概率与用户的时尚趋势和商品的受欢迎程度成正比。
- 每笔交易都用 2025 年的一个随机日期标记时间戳。

这个过程确保了用户行为的多样性, 并反映了不断变化的流行模式。

3.4 语义类别相似性

时尚类别被映射到语义组 (例如, 上衣、下装、外套)。我们手动使用层次化分组模型定义类别间的相似性。相似性函数分配:

- 相同类别的得分为 1.0,
- 在相同语义组中不同项目的得分为 0.8,
- 根据预定义的相似矩阵, 组间组合的分数范围为 0.1–0.4。

这种结构化的相似性使得推荐引擎能够建议不仅在视觉上匹配而且在功能和风格上下文上也相关的物品。例如, 尽管可能存在视觉相似性, 但由于更高的语义一致性, 购买 “开衫” 的用户可能会被推荐 “毛衣” 而不是 “裙子”。

3.5 个性化推荐引擎

推荐引擎通过评估所有候选项与用户的购买历史来为每位用户生成 Top-K 建议。已购买的项目被排除在外, 每个候选项根据从图像内容和行为模拟获得的多个因素进行评分。

评分包含以下指标:

- 视觉相似性: 通过预训练的 CNN 骨干网提取的图像嵌入之间的余弦相似性。
- 流行度得分: 跨所有用户的物品购买动态频率的归一化值。趋势惩罚是一个偶数指数, 通过增加它, 我们可以容忍用户趋势得分和物品流行度之间不是很大的差异。

相关性得分计算如下:

$$\text{Relevance} = w_v \cdot \text{VisSim} + w_p \cdot (1 - (\text{NormPop} - t)^\gamma) + w_c \cdot \text{CatSim}$$

其中:

- w_v 和 w_p 分别是视觉相似性、流行度调节和类别相似性的可调权重。
- t 是用户特定的时尚指数。
- γ 是受欢迎度惩罚指数。
- NormPop 是候选项的极小-极大归一化受欢迎度得分。

候选项目根据最终相关性得分进行排序, 然后将每次购买排名最高的 Top-K 项目作为推荐结果返回。

3.6 地理距离评分

除了视觉和语义上的相关性外, 我们还计算用户位置与每个推荐商品的商店位置之间的物理距离。商店位置被合成地分配了随机的地理坐标, 并使用 Haversine 公式 [9] 计算距离, 具体如下:

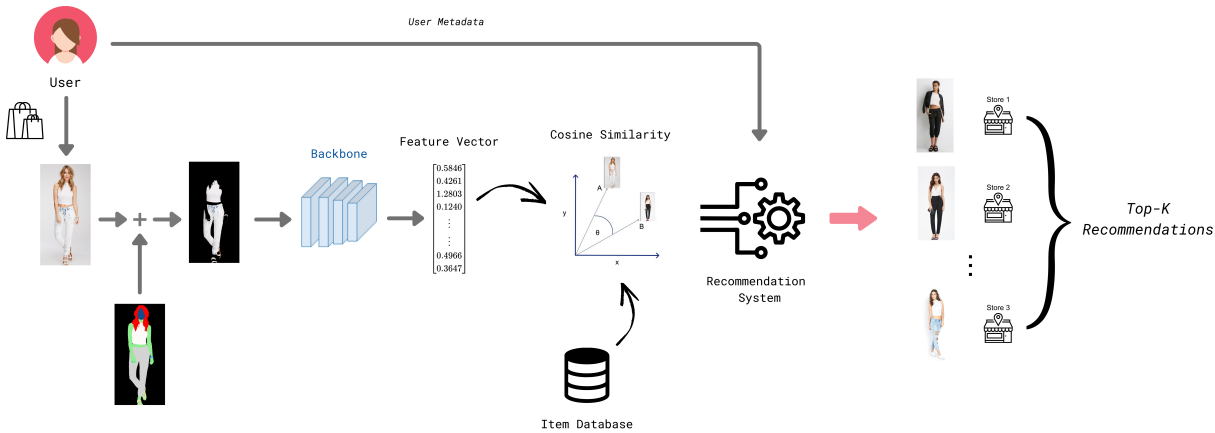


Figure 1: 我们推荐系统的端到端架构，展示了用户输入处理过程，即先前购买的商品、特征提取、余弦相似度计算，以及基于用户元数据生成的前 K 个个性化建议。

$$d = 2r \cdot \arcsin \left(\sqrt{\sin^2 \left(\frac{\Delta \phi}{2} \right) + \cos(\phi_1) \cdot \cos(\phi_2) \cdot \sin^2 \left(\frac{\Delta \lambda}{2} \right)} \right)$$

其中 r 是地球半径（大约 6371 公里）， ϕ 和 λ 是以弧度表示的纬度和经度值， Δ 表示用户和商店坐标之间的差异。虽然目前在排名公式中未使用，但距离信息可用于后续的过滤或可视化，模拟真实世界的用户限制，例如商店的接近性或旅行的努力。

4 实验

4.1 实验设置

为了评估所提出推荐系统的有效性，我们使用 DeepFashion 数据集 [6]，这是一份大型时尚基准，包含超过 80 万张带有类别、身份和属性标签的服装图像。我们提取了一个带有性别和类别标注的项目子集，并使用一致的文件名模式以方便元数据解析。

视觉特征是通过使用三个预训练的 CNN 架构提取的，即 ResNet-50 [3]、DenseNet-121 [4] 和 VGG16 [11]，每个架构都被修改以输出倒数第二层的嵌入。图像被调整为 256×256 大小，然后居中裁剪为 224×224 ，并使用 ImageNet [10] 统计数据进行归一化。最终的视觉特征被缓存以提高效率。

我们模拟了一个为期一年的推荐场景，其中包含 50 个合成用户，每位用户最多可以进行 5 次购买。用户偏好受到个性化流行度评分的影响，该评分调节他们选择热门商品的可能性。每件商品都与五个模拟店铺位置之一相关联，每个位置在二维地理空间中随机分布。

4.2 推荐策略

对于每个用户，我们基于之前购买的商品生成个性化的前 5 推荐（使用 $K = 5$ ）。候选项通过性别兼容性进行过滤，并使用我们之前介绍的加权融合进行评分。我们使用的默认权重配置是 $w_v = 0.7$ ， $w_p = 0.3$ ，但在消融研究中测试了不同的变体。我们选择了受欢迎程度指数为 2。

我们的时尚推荐系统根据视觉相似性和流行度结合的相关性得分对候选项进行排名。为了评估我们基于视觉相似性的时尚推荐系统的有效性，我们关注以下评估标准：

- 类别相似性：衡量推荐项目的类别是否与购买项目的类别接近。
- 性别匹配：衡量推荐商品的性别属性是否与进行原始购买的用户的性别匹配。

为了评估我们基于流行度的时尚推荐系统的有效性。我们使用：

- 流行度 MAE，我们进行归一化，然后计算推荐物品流行度与用户潮流偏好之间的平均绝对误差（MAE）。较小的距离表示更好的效果。

4.3 不同骨干网络的结果

我们评估了不同 CNN 骨干网络——ResNet-50、DenseNet-121 和 VGG16——对推荐中使用的视觉表示质量的影响，以及它如何影响流行度均方误差。每个骨干网络输出一个特征嵌入，用于计算项目之间的余弦相似度。我们在表格 1 中总结了结果。

Backbone	Category Similarity	Gender Alignment	Popularity MAE
ResNet-50 [3]	64.95 %	100 %	12.96
DenseNet-121 [4]	63.53 %	100 %	12.47
VGG16 [11]	57.99 %	100 %	13.32

Table 1: 跨 CNN 主干网络的类别相似性、性别匹配和流行度的 MAE 比较。

所有模型都实现了完美的性别匹配（100 %），这表明仅基于视觉相似度得分对用户性别偏好的建模是可靠的。

DenseNet-121 的类别相似度得分为 63.53 %，ResNet-50 的类别相似度得分为 64.95 %，均优于 VGG16 的 57.99 %，表明其在视觉相似度方面更具优势。受欢迎度 MAE 值变化很小，从 VGG16 的 21.29 到 DenseNet-121 的 21.59 不等。为了区分评分函数中受欢迎度和视觉相似度的贡献，我们进行了一项消融研究。通过将相应权重设为零来禁用评分组件。我们评估去除这些因素对受欢迎度 MAE 和类别相似度的影响。

Model Variant	Category Similarity	Popularity MAE
Full Model	62.16 %	12.92
- No Pop. Modulation	↑ 63.53 %	↑ 31.14
- No Visual Sim.	↓ 40.01 %	↓ 0.22 %

Table 2: 消融研究测量从相关性评分函数中移除组件的效果。

表 2 展示了消融研究的结果，该研究评估了相关性评分函数中不同组件的影响。如预期，去除流行度调制（- 无流行度调制）导致流行度 MAE 显著增加，从 12.92 上升到 31.14。这一配置使得类别相似性略微提高到 63.53 %，相比完整模型的 62.16 %，表明完整模型中使用的权重配置（ $w_v = 0.7, w_p = 0.2$ ）在最大化类别对齐方面接近最佳。相反，禁用视觉相似性（- 无视觉相似性）导致类别相似性下降至 40.01 %，同时将流行度 MAE 减小到仅 0.22。这表明，在没有视觉线索的情况下，模型完全优先考虑流行度信号，导致语义匹配差。这些发现强调了平衡视觉信号和流行度信号以保持个性化和相关推荐的重要性。

4.4 趋势偏好分析

我们研究了用户特定的流行偏好对推荐物品的受欢迎程度和类别相似性的影响。根据用户的流行度得分，用户被分为三组：

- 低潮流性（ $t < 0.33$ ）：偏好于小众或不太流行的物品。
- 中等流行度（ $0.33 \leq t < 0.66$ ）：行为平衡。
- 高趋势性（ $t \geq 0.66$ ）：对流行物品的偏好。

我们计算了每个组的平均受欢迎度 MAE 和类别相似性，以检验用户行为对推荐的影响。我们在表格 3 中展示了结果。

Trendiness Group	Category Similarity	Popularity MAE
Low Trendiness	62.52 %	11.84
Medium Trendiness	65.24 %	13.96
High Trendiness	56.3 %	12.05

Table 3: 用户潮流性对推荐的受欢迎程度和多样性的影响。

表格 3 的结果揭示了用户对趋势偏好和推荐行为之间的细微互动。高趋势得分的用户——那些偏好流行物品的用户——表现出最低的类别相似度为 56.3 %，这表明他们的推荐与他们的类别兴趣的对齐程度较低。相反，中等趋势的用户显示出最高的类别相似度为 65.24 % 和最高的流行度 MAE 为 13.96，这表明一种平衡的推荐策略，即系统虽然更偏离大众流行度，但仍然结合了个人兴趣。低趋势的用户，偏好小众物品，达到相对较高的类别相似度为 62.52 % 和最低的流行度 MAE 为 11.84，这表明模型能够有效地为他们推荐小众但主题相关的物品。

4.5 定性案例研究

为了定性评估推荐引擎的行为，我们展示了一位样本用户的视觉结果。对于用户已购买的一个产品，我们展示了对使用三种基础模型生成的前 5 个推荐项目。此可视化实现了对比基础模型在视觉一致性、风格变异和类别匹配方面的差异。



Figure 3: 在 ResNet-50、DenseNet-121 和 VGG16 下，为一个示例用户提供的前五推荐。

这些可视化结果显示，尽管所有模型都提供了合理的建议，但推荐的特征有所不同。例如，ResNet-50 倾向于生成视觉上相似的装扮，而 VGG16 则引入了更多的风格变化。DenseNet-121 通常在两者之间取得平衡。

在这项工作中，我们通过生成一个合成数据集来模拟一个社会购物环境，该数据集捕捉了受个体时尚偏好影响的用户-项目交互。这一设置使我们能够在可控的环境中探索用户行为与推荐动态之间的相互作用。我们使用定性和定量指标对系统进行了评估，包括类别相似性、性别匹配和受欢迎度距离，以评估推荐与用户/项目特征之间的一致性。

我们的研究表明，所提出的模型在维持性别对齐方面是有效的，并且在平衡类别相关性和受欢迎度方面表现良好，特别是在流行趋势偏好适度表达时。然而，我们观察到当输入为包含多件服装的全身图像时，视觉相似性存在局限性。这表明在应用目标检测以更精确地隔离和比较单个物品方面有潜在的改进，这可以提高类别相似性评分和整体相关性估计。

展望未来，可以将已验证的真实信号（例如性别和类别对齐）直接集成到相关性评分机制中，适当加权以引导模型做出更个性化和情境感知的推荐。

References

- [1] M S Arunkumar, R Gopinath, M Chandru, R Suguna, S Deepa, and V Omprasath. 2024. Fashion Recommendation System for E-Commerce using Deep Learning Algorithms. In *2024 15th International Conference on Computing Communication and Networking Technologies (ICCCNT)*. 1–7. doi:10.1109/ICCCNT61001.2024.10724655
- [2] Guibing Guo. 2012. Resolving data sparsity and cold start in recommender systems. In *User Modeling, Adaptation, and Personalization: 20th International Conference, UMAP 2012, Montreal, Canada, July 16–20, 2012. Proceedings 20*. Springer, 361–364.
- [3] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. 2016. Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition*. 770–778.
- [4] Gao Huang, Zhuang Liu, Laurens Van Der Maaten, and Kilian Q Weinberger. 2017. Densely connected convolutional networks. In *Proceedings of the IEEE conference on computer vision and pattern recognition*. 4700–4708.
- [5] Brett Koonce. 2021. ResNet 50. In *Convolutional neural networks with swift for tensorflow: image recognition and dataset categorization*. Springer, 63–72.
- [6] Ziwei Liu, Ping Luo, Shi Qiu, Xiaogang Wang, and Xiaoou Tang. 2016. Deep-Fashion: Powering Robust Clothes Recognition and Retrieval With Rich Annotations. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*.
- [7] Vivek Raj Singh3 Vineet Swami4 Anushika Tyagi5 Shivam Verma6 NehaVarma1, Garima Singh2. 2024. STYLESYNC: A FASHION RECOMMENDATION SYSTEM. In *2024 International journal of scientific research in Engineering*

- and management.
- [8] Keiron O'shea and Ryan Nash. 2015. An introduction to convolutional neural networks. *arXiv preprint arXiv:1511.08458* (2015).
 - [9] P.N. Ruane. 2014. Heavenly mathematics: the forgotten art of spherical trigonometry by Glen Van Brummelen, pp 192, £24.95, ISBN 978-0-691-14892-2, Princeton University Press (2013). *The Mathematical Gazette* 98 (11 2014), 561–562. doi:10.1017/S0025557200008573
 - [10] Olga Russakovsky, Jia Deng, Hao Su, Jonathan Krause, Sanjeev Satheesh, Sean Ma, Zhiheng Huang, Andrej Karpathy, Aditya Khosla, Michael Bernstein, et al. 2015. Imagenet large scale visual recognition challenge. *International journal of computer vision* 115 (2015), 211–252.
 - [11] Karen Simonyan and Andrew Zisserman. 2014. Very deep convolutional networks for large-scale image recognition. *arXiv preprint arXiv:1409.1556* (2014).
 - [12] Lingala Sivaranjani, Sandeep Kumar Rachamadugu, B.V. Suresh Reddy, Basi Reddy A, M. Sakthivel, and Sivakumar Depuru. 2023. Fashion Recommendation System Using Machine Learning. In *2023 4th International Conference on Smart Electronics and Communication (ICOSEC)*. 1367–1374. doi:10.1109/ICOSEC58147.2023.10275967
 - [13] Bereket A. Yilma, Chan Mi Kim, Gerald C. Cupchik, and Luis A. Leiva. 2024. Artful Path to Healing: Using Machine Learning for Visual Art Recommendation to Prevent and Reduce Post-Intensive Care Syndrome (PICS). In *Proceedings of the 2024 CHI Conference on Human Factors in Computing Systems* (Honolulu, HI, USA) (*CHI '24*). Association for Computing Machinery, New York, NY, USA, Article 447, 19 pages. doi:10.1145/3613904.3642636
 - [14] Bereket A. Yilma and Luis A. Leiva. 2023. The Elements of Visual Art Recommendation: Learning Latent Semantic Representations of Paintings. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems* (Hamburg, Germany) (*CHI '23*). Association for Computing Machinery, New York, NY, USA, Article 24, 17 pages. doi:10.1145/3544548.3581477
 - [15] Bereket A. Yilma and Luis A. Leiva. 2023. Together Yet Apart: Multimodal Representation Learning for Personalised Visual Art Recommendation. In *Proceedings of the 31st ACM Conference on User Modeling, Adaptation and Personalization* (Limassol, Cyprus) (*UMAP '23*). Association for Computing Machinery, New York, NY, USA, 204–214. doi:10.1145/3565472.3592964