

Knowledge-Guided Colorization: Overview, Prospects and Challenges

Rory Ward¹, M. Jaleed Khan¹, John G. Breslin^{1,2} and Edward Curry^{1,2}

¹*SFI Centre for Research Training in Artificial Intelligence, Data Science Institute, University of Galway, Ireland.*

²*Insight SFI Research Centre for Data Analytics, Data Science Institute, University of Galway, Ireland.*

Abstract

Automatic image colorization is notorious for being an ill-posed problem, i.e., multiple plausible colorizations exist for any given black-and-white image. Current approaches to this task revolve around deep neural network-based systems, which do not incorporate knowledge into their colorizations. We present Knowledge-Guided Colorization as a possible solution to the above-mentioned problems. Knowledge-Guided Colorization combines a deep learning-based colorization system and a knowledge graph to inform its colorizations. This is the first time these two techniques have been combined for colorization. The prospects of knowledge-guided colorization are promising, with various potential application scenarios. However, several associated challenges are also highlighted in this research.

Keywords

Colorization, Knowledge Graph, Explainability

1. Introduction

Image colorization is the process of applying color information to black-and-white images. It is an exciting task as many black-and-white photos are lost to history and not getting the exposure to the public that they merit. The modern audience is no longer interested in dull appearing pictures. Image colorization is an ill-constrained task, as multiple color solutions exist for any given black-and-white image. This makes the process difficult as much research must be done to colorize accurately. Automatic image colorization has been proposed to aid in some of this work. Automated image colorization typically relies on some form of a neural network, which requires varying amounts of input from the user. This amount of user involvement ranges from in-depth color hints, like in scribble-based systems [1], to almost no intervention, like in the case of fully automatic image colorization [2].

Machine learning models use Knowledge Graphs (KGs) to incorporate explicit semantics and factual knowledge as a common sense knowledge source, leading to improved model performance and robustness and enhanced reasoning capabilities and interoperability [3].

NeSy'23: International Workshop on Neural-Symbolic Learning and Reasoning, July 03–05, 2023, Siena, Italy

*Corresponding author.

✉ r.ward15@universityofgalway.ie (R. Ward); m.khan12@universityofgalway.ie (M. J. Khan); john.breslin@universityofgalway.ie (J. G. Breslin); edward.curry@universityofgalway.ie (E. Curry)

🌐 <http://www.johnbreslin.com/> (J. G. Breslin); <https://edwardcurry.org/> (E. Curry)

🆔 0009-0003-7634-9946 (R. Ward); 0000-0003-4727-4722 (M. J. Khan); 0000-0001-5790-050X (J. G. Breslin); 0000-0001-8236-6433 (E. Curry)



© 2023 Copyright for this paper by its authors. Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).



CEUR Workshop Proceedings (CEUR-WS.org)

Integrating KGs as a common sense knowledge source in neuro-symbolic visual understanding and reasoning techniques has emerged as a promising research direction [4]. Recent works [5, 6, 7, 8] have demonstrated that image representation and reasoning techniques can effectively capture and interpret detailed semantics in images by utilizing related facts and background knowledge of visual concepts from KGs.

KGs contain valuable information about the colors of objects, the shades in different scenes, and how they vary with diverse backgrounds and conditions. This structured knowledge about colors and shades can serve as a guide for deep learning-based colorization techniques. Although there is no dedicated KG for colorization, general-purpose KGs such as ConceptNet [9], and CSKG [10] contains basic knowledge about colors and color-related attributes of entities. By utilizing this subset of existing KGs, colorization models can leverage the available knowledge to predict the accurate colors of different objects in the image. This approach can significantly improve the quality and realism of the colorized image by incorporating knowledge from external sources beyond the image itself. Furthermore, KG-based colorization techniques can be particularly useful when color information is limited or missing in the original image, enabling the model to make more informed and accurate color predictions.

2. Related Work

When comparing the related work in image colorization, they can generally be clustered by their network architecture. The most common architectures are Convolutional Neural Network (CNN) [11, 12, 13, 14, 15, 16, 17], Generative Adversarial Network (GAN) [18, 19, 20, 21, 22, 23, 24, 25], transformer [26, 27, 28], and diffusion-based [29] systems.

”Real-Time User-Guided Image Colorization with Learned Deep Priors” (Real-Time) [14] is an example of a CNN-based colorization system. It is a fascinating system because it is designed with a user-in-mind methodology. It takes in sparse ”hints” from the user and colorizes the image based on these ”hints”. One of the novel pieces of this work is that they trained this network using randomly simulated user inputs, meaning that the training did not require much human intervention. DeOldify [30] is an example of a GAN-based colorization system. It is based on a ”Self-Attention Generative Adversarial Network” [31] with spectral normalization [32]. It uses a ”Two Time-Scale Update Rule” [33] and NoGAN [34] training. ColTran [26] is an example of a transformer-based colorization system. It is based on self-attention. Initially, it uses a conditional autoregressive transformer to produce a low-resolution coarse coloring of the grayscale image. Then it upsamples the coarse-colored low-resolution image into a finely colored high-resolution image. Palette [29] is an example of a diffusion-based colorization system. It is a generalist image-to-image translation system built on conditional diffusion [35].

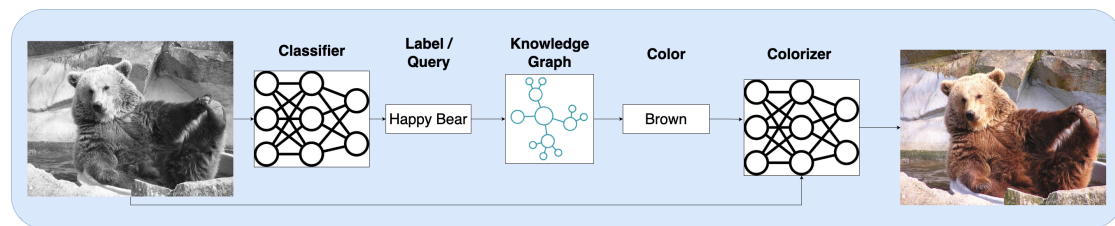
Visual understanding and reasoning techniques have shown great promise in leveraging KGs to extract relevant information and embed it within machine learning models [4]. Graph-based approaches employ message-passing mechanisms to embed structural information from the KG within the model’s representations. For instance, GB-Net [7] links entities and edges in a scene graph to corresponding entities and edges in a common sense graph extracted from Visual Genome [36], WordNet [37], and ConceptNet [9]. The model iteratively refines the scene graph using GNN-based message passing. Similarly, Guo et al. [38] used an instance

relation transformer to extract relational and common sense knowledge from Visual Genome and ConceptNet for scene graph generation. Khan et al. [5] employed a multi-modal deep learning pipeline for scene graph generation, followed by KG-based enrichment to improve the accuracy and expressiveness of image representations. In a different application, Castellano et al. [8] presented ArtGraph, an artistic KG based on WikiArt and DBpedia, and ArtGraph-based fine art classification method for artwork attribute prediction. The technique extracts embeddings from ArtGraph and injects them as external knowledge into a deep learning model, achieving state-of-the-art performance.

3. Knowledge-Guided Colorization

The proposition is to guide colorization using color knowledge extracted from KGs. The concept is that instead of relying on an end-to-end colorizer to learn what color particular objects are, it can be guided using knowledge. The overall system consists of an image classifier, a knowledge graph, and a colorizer, see Fig 1. Initially, the black-and-white image is classified using a neural network. The predicted class is used to query the KG for the given object’s color. From there, this color embedding is passed along with the initial black-and-white image to the colorizer, which produces the final colorized image.

Figure 1: Network Diagram of Knowledge-Guided Colorization. It comprises an image classifier, a knowledge graph, and an image colorizer. This figure shows how the information flows through the system. Image is taken from Visual Genome [36].



This method works equally well for simple and complex scenes. KGs are extremely good for modeling relationships. Let us suppose that in a scene, there is a sun in the sky and no clouds; we can, therefore, deduce that the sky is blue instead of overcast or grey. This would require multiple "hops" over the knowledge graph to supply contextual colorization knowledge.

4. Prospects, Challenges, and Scenarios

4.1. Prospects

The main prospects are that Knowledge-Guided Colorization could improve colorization accuracy, explainability, consistency, and efficiency. **Accuracy:** as objects with multiple plausible colorizations could be colorized faithfully to a KG. **Explainability:** any given colorization could be explained by examining the nodes the system used to guide its colorization. **Consistency:** as the knowledge in the KG would be consistent, unlike the outputs of a neural network.

Efficiency: as fewer training examples would be needed for the system to learn something implicitly that it can be told explicitly. Furthermore, a novel colorization-specific knowledge graph could be produced. This could provide more relevant knowledge than traditional KGs, where only a subset of the information will apply to colorization.

With the semantic understanding of objects, scenes, and their relationships, colorization models can accurately predict the colors of objects or the color palette of a scene based on real-world knowledge. KGs can provide knowledge about the most probable colors based on the object's type or surrounding elements, which can help resolve ambiguities related to object colors in complex scenes. For example, in a grayscale beach scene, leveraging ConceptNet can help assign appropriate colors, i.e., brownish for sand, blue for ocean and sky, and green for palm leaves, resulting in realistic colorization.

4.2. Challenges

There are several challenges associated with leveraging KGs for image colorization. The fundamental challenge is effectively **bridging the gap between both domains**, i.e. the structured, symbolic representation of KGs and the unstructured pixel-level information required for colorization. A major challenge is determining the optimal method for **infusing KG embeddings** into a deep learning-based colorization network. This entails assessing various approaches, such as fusing the knowledge embeddings with the visual embeddings during feature extraction or more advanced techniques like attention mechanisms. The **unavailability of color-specific knowledge graphs** is also a challenge. General KGs, such as ConceptNet [9], contains color information about objects in the world, which can provide basic cues about the typical colors of objects. However, the **contextual correctness** of color knowledge extracted from KGs cannot be guaranteed, as color perception is influenced by factors such as illumination, shadows, and light interactions. The general KGs capture high-level concepts and relationships but do not provide sufficient context to resolve ambiguities in object colors. For example, an object like a car can have various colors, but ConceptNet might not offer specific guidance on which color to use in a particular setting.

4.3. Scenarios

Knowledge-Guided Colorization could be applicable in several diverse scenarios. One of the significant interest areas would be in the realm of historical images. Generally, there is little data available for objects in historical images, such as military uniforms from a particular era. This typically leads to inaccurate colorizations from traditional end-to-end colorizers. This is where adding knowledge from a KG can enable more accurate colorizations. Colors can be essential in historic colorizations as a change in color can change the whole story of an image, for example, which wars a particular decorated soldier fought based on their medals.

Acknowledgments

This work was conducted with the financial support of the Science Foundation Ireland Centre for Research Training in Artificial Intelligence under Grant No.18/CRT/6223, and also by a grant

from Science Foundation Ireland under Grant Number SFI/12/RC/2289_P2. For the purpose of Open Access, the author has applied a CC BY public copyright licence to any Author Accepted Manuscript version arising from this submission.

References

- [1] A. Levin, D. Lischinski, Y. Weiss, Colorization using optimization, *ACM Transactions on Graphics* 23 (2004). doi:10.1145/1015706.1015780.
- [2] D. Varga, T. Szirányi, Fully automatic image colorization based on convolutional neural network, in: *2016 23rd International Conference on Pattern Recognition (ICPR)*, 2016, pp. 3691–3696. doi:10.1109/ICPR.2016.7900208.
- [3] U. Kursuncu, M. Gaur, A. Sheth, Knowledge infused learning (k-il): Towards deep incorporation of knowledge in deep learning, *arXiv preprint arXiv:1912.00512* (2019).
- [4] M. J. Khan, J. G. Breslin, E. Curry, Common sense knowledge infusion for visual understanding and reasoning: Approaches, challenges, and applications, *IEEE Internet Computing* 26 (2022) 21–27.
- [5] M. J. Khan, J. G. Breslin, E. Curry, Expressive scene graph generation using commonsense knowledge infusion for visual understanding and reasoning, in: *The Semantic Web: 19th International Conference, ESWC 2022, Hersonissos, Crete, Greece, May 29–June 2, 2022, Proceedings*, Springer, 2022, pp. 93–112.
- [6] Z. Chen, J. Chen, Y. Geng, J. Z. Pan, Z. Yuan, H. Chen, Zero-shot visual question answering using knowledge graph, in: *The Semantic Web–ISWC 2021: 20th International Semantic Web Conference, ISWC 2021, Virtual Event, October 24–28, 2021, Proceedings 20*, Springer, 2021, pp. 146–162.
- [7] A. Zareian, S. Karaman, S.-F. Chang, Bridging knowledge graphs to generate scene graphs, in: *Computer Vision–ECCV 2020: 16th European Conference, Glasgow, UK, August 23–28, 2020, Proceedings, Part XXIII 16*, Springer, 2020, pp. 606–623.
- [8] G. Castellano, V. Digeno, G. Sansaro, G. Vessio, Leveraging knowledge graphs and deep learning for automatic art analysis, *Knowledge-Based Systems* 248 (2022) 108859.
- [9] R. Speer, J. Chin, C. Havasi, Conceptnet 5.5: An open multilingual graph of general knowledge, in: *Proceedings of the AAAI conference on artificial intelligence*, volume 31, 2017.
- [10] F. Ilievski, P. Szekely, B. Zhang, Cskg: The commonsense knowledge graph, in: *The Semantic Web: 18th International Conference, ESWC 2021, Virtual Event, June 6–10, 2021, Proceedings 18*, Springer, 2021, pp. 680–696.
- [11] S. Iizuka, E. Simo-Serra, DeepRemaster: Temporal Source-Reference Attention Networks for Comprehensive Video Enhancement, *ACM Transactions on Graphics (Proc. of SIGGRAPH Asia 2019)* 38 (2019) 1–13.
- [12] G. Larsson, M. Maire, G. Shakhnarovich, Learning representations for automatic colorization, 2016. URL: <https://arxiv.org/abs/1603.06668>. doi:10.48550/ARXIV.1603.06668.
- [13] M. He, D. Chen, J. Liao, P. V. Sander, L. Yuan, Deep exemplar-based colorization, 2018. URL: <https://arxiv.org/abs/1807.06587>. doi:10.48550/ARXIV.1807.06587.
- [14] R. Zhang, J.-Y. Zhu, P. Isola, X. Geng, A. S. Lin, T. Yu, A. A. Efros, Real-time user-guided

- image colorization with learned deep priors, 2017. URL: <https://arxiv.org/abs/1705.02999>. doi:10.48550/ARXIV.1705.02999.
- [15] J.-W. Su, H.-K. Chu, J.-B. Huang, Instance-aware image colorization, 2020. URL: <https://arxiv.org/abs/2005.10825>. doi:10.48550/ARXIV.2005.10825.
- [16] R. Zhang, P. Isola, A. A. Efros, Colorful image colorization, 2016. URL: <https://arxiv.org/abs/1603.08511>. doi:10.48550/ARXIV.1603.08511.
- [17] M. Dias, J. Monteiro, J. Estima, J. Silva, B. Martins, Semantic segmentation and colorization of grayscale aerial imagery with w-net models, *Expert Systems* 37 (2020) e12622. URL: <https://onlinelibrary.wiley.com/doi/abs/10.1111/exsy.12622>. doi:<https://doi.org/10.1111/exsy.12622>. arXiv:<https://onlinelibrary.wiley.com/doi/pdf/10.1111/exsy.12622>.
- [18] P. Isola, J.-Y. Zhu, T. Zhou, A. A. Efros, Image-to-image translation with conditional adversarial networks, 2016. URL: <https://arxiv.org/abs/1611.07004>. doi:10.48550/ARXIV.1611.07004.
- [19] C. Zou, H. Mo, C. Gao, R. Du, H. Fu, Language-based colorization of scene sketches, *ACM Trans. Graph.* 38 (2019). URL: <https://doi.org/10.1145/3355089.3356561>. doi:10.1145/3355089.3356561.
- [20] L. Zhang, C. Li, T.-T. Wong, Y. Ji, C. Liu, Two-stage sketch colorization, *ACM Trans. Graph.* 37 (2018). URL: <https://doi.org/10.1145/3272127.3275090>. doi:10.1145/3272127.3275090.
- [21] X. Kuang, X. Sui, C. Liu, Y. Liu, Q. Chen, G. Gu, Thermal infrared colorization via conditional generative adversarial network, 2018. URL: <https://arxiv.org/abs/1810.05399>. doi:10.48550/ARXIV.1810.05399.
- [22] W. Chen, J. Hays, Sketchygan: Towards diverse and realistic sketch to image synthesis, 2018. URL: <https://arxiv.org/abs/1801.02753>. doi:10.48550/ARXIV.1801.02753.
- [23] P. Hensman, K. Aizawa, cgan-based manga colorization using a single training image, 2017. URL: <https://arxiv.org/abs/1706.06918>. doi:10.48550/ARXIV.1706.06918.
- [24] C. W. Seo, Y. Seo, Seg2pix: Few shot training line art colorization with segmented image data, *Applied Sciences* 11 (2021) 1464. URL: <http://dx.doi.org/10.3390/app11041464>. doi:10.3390/app11041464.
- [25] Y. Cao, Z. Zhou, W. Zhang, Y. Yu, Unsupervised diverse colorization via generative adversarial networks, 2017. URL: <https://arxiv.org/abs/1702.06674>. doi:10.48550/ARXIV.1702.06674.
- [26] M. Kumar, D. Weissenborn, N. Kalchbrenner, Colorization transformer, 2021. URL: <https://arxiv.org/abs/2102.04432>. doi:10.48550/ARXIV.2102.04432.
- [27] E. Casey, V. Pérez, Z. Li, H. Teitelman, N. Boyajian, T. Pulver, M. Manh, W. Grisaitis, The animation transformer: Visual correspondence via segment matching, *CoRR* abs/2109.02614 (2021). URL: <https://arxiv.org/abs/2109.02614>. arXiv:2109.02614.
- [28] Z. Wan, B. Zhang, D. Chen, J. Liao, Bringing old films back to life, 2022. URL: <https://arxiv.org/abs/2203.17276>. doi:10.48550/ARXIV.2203.17276.
- [29] C. Saharia, W. Chan, H. Chang, C. A. Lee, J. Ho, T. Salimans, D. J. Fleet, M. Norouzi, Palette: Image-to-image diffusion models, 2021. URL: <https://arxiv.org/abs/2111.05826>. doi:10.48550/ARXIV.2111.05826.
- [30] J. Antic, Deoldify, <https://github.com/jantic/DeOldify>, 2019.
- [31] H. Zhang, I. Goodfellow, D. Metaxas, A. Odena, Self-attention generative adversarial

- networks, 2018. URL: <https://arxiv.org/abs/1805.08318>. doi:10.48550/ARXIV.1805.08318.
- [32] T. Miyato, T. Kataoka, M. Koyama, Y. Yoshida, Spectral normalization for generative adversarial networks, 2018. URL: <https://arxiv.org/abs/1802.05957>. doi:10.48550/ARXIV.1802.05957.
- [33] M. Heusel, H. Ramsauer, T. Unterthiner, B. Nessler, S. Hochreiter, Gans trained by a two time-scale update rule converge to a local nash equilibrium (2017). URL: <https://arxiv.org/abs/1706.08500>. doi:10.48550/ARXIV.1706.08500.
- [34] J. H. f. Jason Antic (Deoldify), Decrappification, deoldification, and super resolution, 2019. URL: <https://www.fast.ai/posts/2019-05-03-decrappify.html>.
- [35] J. Sohl-Dickstein, E. Weiss, N. Maheswaranathan, S. Ganguli, Deep unsupervised learning using nonequilibrium thermodynamics, in: International Conference on Machine Learning, PMLR, 2015, pp. 2256–2265.
- [36] R. Krishna, Y. Zhu, O. Groth, J. Johnson, K. Hata, J. Kravitz, S. Chen, Y. Kalanditis, L.-J. Li, D. A. Shamma, M. Bernstein, L. Fei-Fei, Visual genome: Connecting language and vision using crowdsourced dense image annotations, 2016.
- [37] G. A. Miller, R. Beckwith, C. Fellbaum, D. Gross, K. J. Miller, Introduction to wordnet: An on-line lexical database, International journal of lexicography 3 (1990) 235–244.
- [38] Y. Guo, J. Song, L. Gao, H. T. Shen, One-shot scene graph generation, in: Proceedings of the 28th ACM International Conference on Multimedia, 2020, pp. 3090–3098.