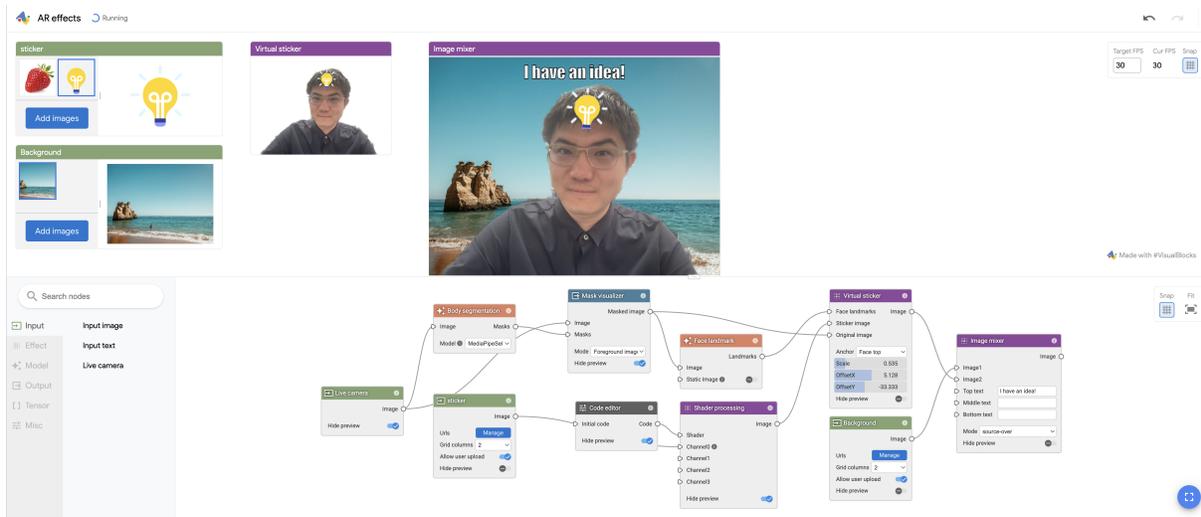# Experiencing Visual Blocks for ML:
Visual Prototyping of AI Pipelines

Ruofei Du, Na Li, Jing Jin, Michelle Carney, Xiuxiu Yuan, Kristen Wright
Mark Sherwood, Jason Mayes, Lin Chen, Jun Jiang, Jingtao Zhou, Zhongyi Zhou
Ping Yu, Adarsh Kowdle, Ram Iyengar, and Alex Olwal
Contact:me@duruofei.com,linazhao@google.com
Google Research

Figure 1: Visual Blocks for ML is a rapid visual prototyping system that allows users to build and share ML pipelines. In this example, the user builds a real-time pipeline that depicts an AR sticker on the user' head with a virtual background.

ABSTRACT

We demonstrate Visual Blocks for ML, a visual programming platform that facilitates rapid prototyping of ML-based multimedia applications. As the public version of Rapsai [3], we further integrated large language models and custom APIs into the platform. In this demonstration, we will showcase how to build interactive AI pipelines in a few drag-and-drops, how to perform interactive data augmentation, and how to integrate pipelines into Colabs. In addition, we demonstrate a wide range of community-contributed pipelines in Visual Blocks for ML, covering various aspects including interactive graphics, chains of large language models, computer vision, and multi-modal applications. Finally, we encourage students, designers, and ML practitioners to contribute ML pipelines through https://github.com/google/visualblocks/tree/main/pipelines to inspire creative use cases. Visual Blocks for ML is available at http://visualblocks.withgoogle.com.

CCS CONCEPTS

• **Computing methodologies** → Visual analytics; *Machine learning*; • **Software and its engineering** → **Visual languages**.

KEYWORDS

visual programming, large language models, visual prototyping, multi-modal models, node-graph editor, deep neural networks, data augmentation, deep learning, visual analytics

ACM Reference Format:
Ruofei Du, Na Li, Jing Jin, Michelle Carney, Xiuxiu Yuan, Kristen Wright, Mark Sherwood, Jason Mayes, Lin Chen, Jun Jiang, Jingtao Zhou, Zhongyi Zhou, and Ping Yu, Adarsh Kowdle, Ram Iyengar, and Alex Olwal. 2023. Experiencing Visual Blocks for ML: Visual Prototyping of AI Pipelines. In *The 36th Annual ACM Symposium on User Interface Software and Technology (UIST '23 Adjunct), October 29-November 1, 2023, San Francisco, CA, USA.* ACM, New York, NY, USA

1 INTRODUCTION

Although user-friendly tools for developing machine learning (ML) models in language [2, 7–9] and image classification [1, 4–6] have become readily available, there remains a significant gap in tools supporting real-time multimedia applications. The current tools are

inadequate in efficiently handling visual and audio data from real-world sources, such as camera streams, and fail to enable interactive experimentation with data augmentation and model comparison.

Furthermore, developing and iterating on these ML-based multimedia prototypes can be challenging and costly. It usually involves a cross-functional team of ML practitioners who fine-tune the models, evaluate robustness, characterize strengths and weaknesses, inspect performance in the end-use context, and develop the applications. Moreover, models are frequently updated and require repeated integration efforts before evaluation can occur, which makes the workflow ill-suited to design and experiment.

In this demonstration paper, we present Visual Blocks for ML, an iterated system and public version of Rapsai [3]. Visual Blocks for ML uses a node-graph editor that facilitates rapid prototyping of ML-based multimedia applications. Users can create and connect different components (nodes) to rapidly build an ML pipeline, and see the results in real-time without writing any code. We demonstrate how this platform enables a better model evaluation experience through interactive characterization and visualization of ML model performance and interactive data augmentation and comparison. We have released the Visual Blocks for ML framework, along with a demo and Colab examples [1].

2 SYSTEM OVERVIEW

We iteratively developed Visual Blocks for ML on the base of *Rapsai* [3], a web-based rapid prototyping platform we built where researchers and developers can quickly build and deploy multimedia pipelines. In additional to the initial version, we added support of large language models, multi-modal models, and custom APIs to facilitate creating more diverse and creative pipelines.

2.1 Visual Blocks for ML

As shown in Figure 1, Visual Blocks for ML consists of three coordinated panels: (a) Preview Panel, (b) Nodes Library, and (c) Node-graph Editor. The preview panel depicts visible nodes, where users could upload images, change input text, and preview intermediate steps. The nodes library categorizes 38 available nodes into input, effects, output, model, tensor, and misc nodes. Taking feedback from [3], we integrate node parameters within the node-graph editor, so that users could interactive augment data, change parameters in an intuitive way.

2.2 Visual Blocks for ML Colaboratory

Visual Blocks for ML also supports Google Colab[2]. On the Colab, the user can easily customized their node and serve their web-page on the web interface. The Colab will automatically render the Visual Blocks for ML interface for creative uses.

2.3 Use Cases and Scenarios

The major use scenarios of Visual Blocks for ML can be categorized by the types of ML pipeline made by the user.

Language Pipelines: Our community users have built various language-based pipeline, including "email tone explorer", "weather

summarizer" and even the interactions in the were-wolf game, a role-playing board game, using LLMs as artificial agents.

Vision Pipelines: With Visual Blocks for ML, developers can interactively evaluate the robustness of their computer vision models. For example, the user can add an image processor before an object detection model to test whether the detection can be robust to various lighting conditions.

Multi-modal Pipelines: We have incorporated text-to-image generative models (i.e., Imagen) and PaLM 2, a multi-modal LLMs, into Visual Blocks for ML. The user can easily build a visual captioning tool to enhance accessibility and a draft-to-HTML to support creative processes.

3 CONCLUSION

In this demonstration, we presented Visual Blocks for ML, a system designed to lowers development barriers for ML-based multimedia applications. It empowers users to experiment without worrying about coding or technical details. It also facilitates collaboration between designers and developers by providing a common language for describing ML pipelines. In the future, we plan to open this framework up for the community to contribute their own nodes and integrate it into many different platforms. We expect visual programming for machine learning to be a common interface across ML tooling going forward.

REFERENCES

[1] Michelle Carney, Barron Webster, Irene Alvarado, Kyle Phillips, Noura Howell, Jordan Griffith, Jonas Jongejan, Amit Pitaru, and Alexander Chen. 2020. Teachable Machine: Approachable Web-Based Tool for Exploring Machine Learning Classification. In *Extended Abstracts of the 2020 CHI Conference on Human Factors in Computing Systems*. ACM. https://doi.org/10.1145/3334480.3382839

[2] John Joon Young Chung, Wooseok Kim, Kang Min Yoo, Hwaran Lee, Eytan Adar, and Minsuk Chang. 2022. TaleBrush: Sketching Stories With Generative Pretrained Language Models. In *CHI Conference on Human Factors in Computing Systems*. 1–19. https://doi.org/10.1145/3491102.3501819

[3] Ruofei Du, Na Li, Jing Jin, Michelle Carney, Scott Miles, Maria Kleiner, Xiuxiu Yuan, Yinda Zhang, Anuva Kulkarni, Xingyu Liu, Ahmed Sabie, Sergio Escolano, Abhishek Kar, Ping Yu, Ram Iyengar, Adarsh Kowdle, and Alex Olwal. 2023. Rapsai: Accelerating Machine Learning Prototyping of Multimedia Applications Through Visual Programming. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems (CHI)*. ACM. https://doi.org/10.1145/3544548.3581338

[4] Michael Gleicher, Aditya Barve, Xinyi Yu, and Florian Heimerl. 2020. Boxer: Interactive Comparison of Classifier Results. *Computer Graphics Forum* (Jun. 2020). https://doi.org/10.1111/cgf.13972

[5] Marco Tulio Ribeiro, Sameer Singh, and Carlos Guestrin. 2016. "Why Should I Trust You?": Explaining the Predictions of Any Classifier. In *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* (San Francisco, California, USA) *(KDD '16)*. Association for Computing Machinery, New York, NY, USA, 1135–1144. https://doi.org/10.1145/2939672.2939778

[6] Thilo Spinner, Udo Schlegel, Hanna Schafer, and Mennatallah El-Assady. 2019. ExplAIner: a Visual Analytics Framework for Interactive and Explainable Machine Learning. *IEEE Transactions on Visualization and Computer Graphics* (2019). https://doi.org/10.1109/TVCG.2019.2934629

[7] Bingyuan Wu and Yongxiong Wang. 2022. Rich Global Feature Guided Network for Monocular Depth Estimation. *SSRN Electronic Journal* (2022). https://doi.org/10.2139/ssrn.4057946

---

[1]Visual Blocks for ML: http://visualblocks.withgoogle.com
[2]Visual Blocks Quick Start for Colab: https://colab.research.google.com/github/google/visualblocks/blob/main/examples/quick_start_cartoonization.ipynb

[8] Tongshuang Wu, Ellen Jiang, Aaron Donsbach, Jeff Gray, Alejandra Molina, Michael Terry, and Carrie Cai. 2022. PromptChainer: Chaining Large Language Model Prompts Through Visual Programming. In *CHI Conference on Human Factors in Computing Systems Extended Abstracts*. ACM. https://doi.org/10.1145/3491101.3519729

[9] Tongshuang Wu, Michael Terry, and Carrie Cai. 2022. AI Chains: Transparent and Controllable Human-AI Interaction by Chaining Large Language Model Prompts. In *CHI Conference on Human Factors in Computing Systems*. ACM. https://doi.org/10.1145/3491102.3517582