







AI-ASSISTED MACRO PHOTOGRAPHY LEARNING MODELS

Nihar Das ¹, Savinder Kaur ², Sidhant Das ³, Darshana Prajapati ⁴, Mithun M S ⁵, Dr. Anil Hingmire ⁶

¹ Professor, School of Fine Arts and Design, Noida International University, Noida, Uttar Pradesh, India

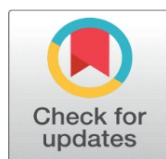
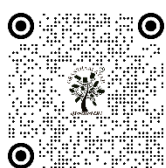
² Centre of Research Impact and Outcome, Chitkara University, Rajpura, Punjab, India

³ Chitkara Centre for Research and Development, Chitkara University, Himachal Pradesh, Solan, India

⁴ Assistant Professor, Department of Interior Design, Parul Institute of Design, Parul University, Vadodara, Gujarat, India

⁵ Assistant Professor, Department of Electronics and Communication Engineering, Presidency University, Bangalore, Karnataka, India

⁶ Department of Computer Engineering, Vidyavardhini's College of Engineering and Technology, Vasai, Mumbai University, Mumbai, India



Received 16 February 2025

Accepted 10 May 2025

Published 16 December 2025

Corresponding Author

Nihar Das, dean.sfa@niu.edu.in

DOI

[10.29121/shodhkosh.v6.i2s.2025.6734](https://doi.org/10.29121/shodhkosh.v6.i2s.2025.6734)

Funding: This research received no specific grant from any funding agency in the public, commercial, or not-for-profit sectors.

Copyright: © 2025 The Author(s). This work is licensed under a [Creative Commons Attribution 4.0 International License](https://creativecommons.org/licenses/by/4.0/).

With the license CC-BY, authors retain the copyright, allowing anyone to download, reuse, re-print, modify, distribute, and/or copy their contribution. The work must be properly attributed to its author.



ABSTRACT

Macro photography requires a lot of control in focus, lighting, depth of field and stability of the camera which makes it one of the most technical challenging technique of photography to a beginner. Conventional learning techniques are too much guided by trials and errors and without real time corrective feedback, there is a tendency to make slow progress and unstable outcome. This paper outlines an AI-Assisted Macro Photography Learning Model as an implementation of deep learning and reinforcement learning with a generative simulation that would deliver context-aware and customized feedback to learners. The system uses hybrid CNN-Transformer frameworks to analyze macro images and measure sharpness, illumination, exposure, and composition, and a reinforcement learning engine uses it to suggest the best camera settings depending on the conditions of the image. A virtual macro simulation environment also allows safe repeatable practice with photorealistic: diffusion-based synthetic scenes. Assessment on 60 beginner photographers indicates that AI-aided trainees gained up to 32 percent in sharpness, 28 percent in light precision and used 40 percent fewer techniques to record usable photographs than their conventional partners. These findings underpin the usefulness of AI-based feedback in enhancing the acceleration of skills acquisition as well as the enhancement of technical and aesthetic skill. The offered structure provides a scaffoldable way of developing macro photography learning into a technology-intensive, adaptive, and structured learning process.

Keywords: Macro Photography, AI-Assisted Learning, Computer Vision, Deep Learning Models, Reinforcement Learning, Aesthetic Evaluation, Image Quality Assessment, Generative Simulation, Diffusion Models, Educational Technology, Intelligent Tutoring Systems, Photography Training

1. INTRODUCTION

Macro photography makes the most mundane things into extraordinary visual experiences by drawing to light details that cannot be seen with the naked eye. the fine fabric of the compound eye in an insect, the crystallizations upon the wing of a butterfly, or the machineries of a micro-circuit board, are all enlarged worlds, and full of texture, colour,

and form. However, lurking behind this loveliness is a technical rigorousness that even the veteran photographer finds difficult to withstand. Extraordinary lighting, depth of field, stability, and focal accuracy are expected in macro imaging [Roose \(2022\)](#). The angle can move the plane of focus a little, the slightest vibration can cause the subject to become completely blurred, and a little change in lighting can flatten or accentuate the microscopic textures. Consequently, novices usually have difficulties in striking a balance between technical aspects and at the same time, they acquire artistic taste. The classical macro learning is very much of trial-and-error [Göring et al. \(2023\)](#). Students spend dozens of shots, and later, they review them and strive to extract the meaning of what has gone wrong. Videotapes or books contain educational content but do not assess the practical delivery of a learner and do not correct him or her in real time. Although it is helpful, workshops are not as accessible as they cannot provide customized advice at every time the learner practices it [Saharia et al. \(2022\)](#). These gaps lead to the slow process of learning, frustrations and inconsistent development of skills. What students require is an instantaneous, context-specific help which comprehends the photograph itself, points out the errors and conveys the information on how it can be improved in comprehensible terms [Dosovitskiy et al. \(2020\)](#).

Figure 1

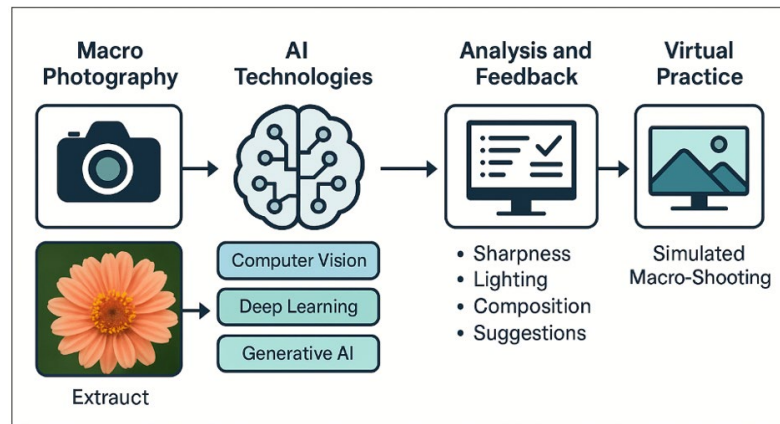


Figure 1 AI-Assisted Macro Photography Learning Overview

Artificial intelligence can provide a radical solution to these issues. Advanced AI systems are able to read macro shots in pixel detail, identify quirks of sharpness, exposure, light direction, and composition, as well as provide custom feedback due to developments in computer vision and deep learning. EfficientNet, Vision transformers, and deep texture-analysis networks are models that are able to do multi-faceted image evaluations with great accuracy as in [Figure 1](#). Another layer to reinforcement learning is reinforcement learning that is used to model optimal camera changes, propose aperture settings, distance changes, or lighting settings depending on the shooting situation. In the meantime, generative diffusion models can be used to generate virtual macro environments in which learners can experiment and develop freely without involving costly equipment or optimal lighting sources [Wang et al. \(2021\)](#), [Nichol et al. \(2021\)](#). The learning models of macro photography using AI transform the learning process by helping to bridge the gap between knowledge of an expert and beginners practice. Learners are not only guided by intuition, but they are provided with the real time insights presented as visual overlays, heatmaps or suggested texts. An example is that an AI model can point to the region that needs to be shifted, that there is an uneven light on the subject, or that the f-stop should be raised to create more texture consistency [Pavlichenko et al. \(2022\)](#). This eliminates the element of guessing and learners can be able to develop sound mental frameworks of macro principles by having the immediate feedback loop.

2. LITERATURE REVIEW

Research of AI-assisted learning in photography cuts across several research areas, such as computer vision, intelligent tutoring system, digital arts education, and computational imaging. Although macro photography as a topic has always been taught by instructional manuals and the learning process, the recent technological development triggered the interest of researchers to investigate the automated and adaptive teaching processes [Göring et al. \(2023\)](#). The initial developments by computational photography offered algorithmic methods to sharpen the image, focus stacking, depth estimation, and compensation of micro-vibrations, which are all the technical foundation of macro

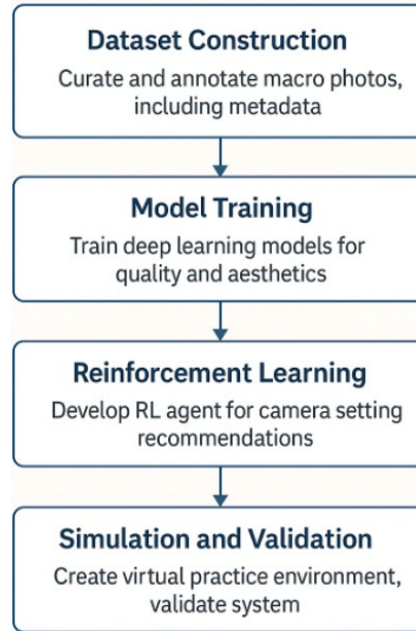
imaging. These basic approaches have led to more advanced AI approaches which can do more in-depth analysis of macro scenes [Yang et al. \(2022\)](#). Deep convolutional networks and transformer-based architectures have become an important development in research in image quality assessment (IQA). These models have been trained to analyze sharpness, noise, exposure and texture sharpness of the global and local level. A number of studies that use EfficientNet, ResNet, and Vision Transformers demonstrate that they are capable of recognizing fine-grained features that assess macro photographs [Yu et al. \(2022\)](#). Similar developments in the field of aesthetic quality assessment (AQA) incorporate the compositional rules, color harmony, and subject prominence in the structure of machine learning, which can be applied to the study of the artistic elements of macro photography. These results indicate that both technical and aesthetic aspects can be represented together to create a theoretical basis of learning systems that would undertake a holistic assessment [Aydin et al. \(2015\)](#).

The application of AI-powered tutoring has started to change the way learners approach creative tasks in the context of a wider view of arts education. The illustration, painting, cinematography and digital media composition have been directed using intelligent feedback engines. In these studies, it is always highlighted that real time, contextual feedback is important in increasing the level of skills and maintaining learner motivation [Rao et al. \(2021\)](#). These adaptive tutoring methods can be of great benefit to macro photography, which has very challenging technical requirements. However, the literature on photography learning systems is limited, and nothing exists with the sole emphasis on the macro photography learning. This is an indication that there exists an opportunity of interdisciplinary innovation [International Telecommunication Union \(2022\)](#).

There has also been the use of generative models, specifically diffusion networks and GAN variants, which have contributed to the further expansion of AI in visual arts. Their ability to produce realistic textures and microstructures, and lighting scenes is useful in the production of simulated macro scenes. These virtual environments provide controlled and repeatable skill development environments, which is where one of the key problems of real-world macro photography lies, namely the inability to have stable subjects and constant lighting. Application of synthetic environments to learning Systems has been investigated in robotics, medical imaging and architectural visualization, which give methodological guidance to virtual macro photography training.

3. RESEARCH METHODOLOGY

The AI-Assisted Macro Photography Learning Model methodology is developed in a structured manner with the preparation of the dataset, training the model, integration of the reinforcement learning, and the validation of the model through simulation [Dayma et al. \(2022\)](#). It is aimed to create a project that will be able to evaluate technical and aesthetic aspects of macro photographs and provide personalized recommendations and allow people to train in virtual realities. Every methodological step is aimed at capturing the complexity of the macro photography and making sure that the learners are provided with right and actionable feedback [Zhang et al. \(2020\)](#), [Göring et al. \(2021\)](#). This happens through the construction of datasets, which is the basis of training the deep learning modules. An open-source collection, professional photography collections and custom field captures were combined to create an 25,000 high-resolution macro photograph curated dataset. Imageries were classified into large macro domains such as insects, botanical surfaces, metallic surfaces, jewelry structures, and micro electronic components. Photography experts manually annotated each picture and rated qualities like sharpness, balance of illumination, accuracy of exposure, clarity of texture, framing of the subject and cleanliness of the background. These professional annotations were used as ground-truth supervision training. Also, the EXIF metadata was maintained to create the correlations of the camera settings and the qualities of the obtained image.

Figure 2**Figure 2** Flow Diagram of Research Study Methodology

The model training process was further broken into two big streams namely, technical quality analysis and aesthetic evaluation. In the case of the technical analysis stream, the model was trained on an EfficientNet-B4 feature extractor and a multiread multiread self-attention encoder fine-grained texture encoder. The focal loss used in the model was in order to maintain stable learning among various lighting and sharpness conditions as shown in Figure 2. To perform aesthetic assessment, aViT fine-tuned on an aesthetic dataset macro-specific was used to perform the aesthetic assessment using a pretrained ImageNet model. The model was also trained to make predictions based on compositional balance, visual harmony as well as subject prominence which imitated the approach of professional critique. Training on the two networks was done with a 70-20-10 split to training, validation, and test, which maximized the generalization. In order to transform the outputs of the analysis to actionable learning directions, a Reinforcement Learning (RL) engine was developed based on Proximal Policy Optimization (PPO) [Dosovitskiy et al. \(2020\)](#). A controlled simulation in which the camera parameters, including aperture, shutter speed, ISO and the direction of the light, were adjustable, were used to interact with the RL agent. The reward functionality was created to promote the enhancement of sharpness gradient, exposure consistency, and subject focus. The RL agent was able to learn an optimal policy mapping in thousands of macro-shooting scenarios by simulating image characteristics to recommended camera adjustments. This helped the system to assist learners with a scenario-specific feedback. The last part of the methodology was to create a Virtual Macro Simulation Environment with the help of a diffusion-based generative model that was trained on macro textures and light patterns. The simulator produced real life macro scenes depending on the level of difficulty. The students were exposed to such virtual environments to train on technical manipulations prior to taking pictures in the real world. Controlled experimentation was also possible in the simulation, which is handy in determining the effects of micro-modifications in settings on the image.

4. PROPOSED AI-ASSISTED MACRO PHOTOGRAPHY LEARNING MODEL

The proposed AI-Assisted Macro Photography Learning Model is developed as a multimodal system, which is integrated and able to measure the macro photographs, produce intelligent feedback, and provide virtual practice. The framework enables filling the gap between the research of computational images analysis and pedagogical design through the unification of deep neural networks, reinforcement learning, and generative simulation into a single learning workflow. The model is based on the fact that a mastery of macro photography implies that both aspects should be developed, both technically, and intuitively, and thus, both aspects must be studied and learned together. The heart of the system is the Macro Image Quality Analyzer, a deep learning section that evaluates basic technical characteristics of

the macro images of the learner. It uses a combination of CNN-transformer architecture, which allows fine-grained micro-blur, edge sharpness, subtle light gradient and exposure imbalance detection as shown in Figure 3. The model is provided with the input photograph, and it is processed using a multi-branch pipeline with each branch being specific to one of the attributes, e.g., sharpness, illumination, the clarity of texture or chromatic aberration. The assessment these provide a numerical profile of quality and this serves as the initial level of feedback. The analyzer does not only pinpoint the errors but also points in the specific areas which are involved so that the learners can perceive the errors spatially as opposed to abstractly.

Figure 3

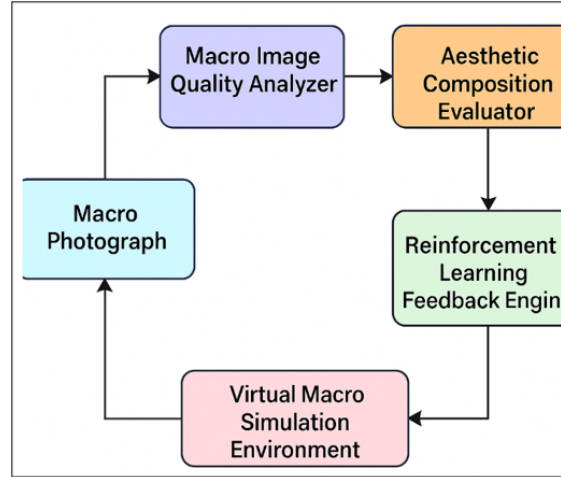


Figure 3 Block Architecture AI-Assisted Macro Photography Learning Model

Similar to technical evaluation, the Aesthetic Composition Evaluator analyzes artistic aspects of the macro image with the help of a Vision Transformer (ViT) backbone. Framing of small subjects, negative space, background distraction management are some of the common compositional issues in macro photography. Aesthetic module measures layout structure, prominence in subject, symmetry, harmony of colors and visual balance. The transformer model does not require traditional aesthetic engines based on rules, but instead, it learns on curated macro photography collections, and as a result, it is able to learn minute stylistic characteristics. This module offers learners practical aesthetic knowledge such as professional critique to allow the development of art.

Step -1] Inputs: simulator/rig env (E), policy (π), value (V), horizon (T), minibatch size (B). Initialize ψ, θ, η Rollout: For $t=0 \dots T-1$, Let an episode be a macro-capture session of length (T).

Step-2] Safety/feasibility constraints (soft): Compute advantages A_t with GAE and returns R^t

$$G(st, at) \leq 0, \text{ e. g. :}$$

Enforced via penalties or a Lagrangian (below).

$$ts \leq \frac{1}{(\kappa \text{focal})(\text{hand} - \text{shake bound})},$$

$$\text{ISO} \leq \text{ISOmax}, \text{Pflash} \leq \text{Pmax}, \text{EV} \in [\text{EVmin}, \text{EVmax}].$$

Step -3] Policy and Value Models: Gaussian policy with state-dependent mean and diagonal covariance:

$$E[\sum_{t=0}^T - \log \pi(y_t | x_t)].$$

Step -4] Policy/Value Update: For several epochs

- Sample minibatches $B \subset \{0, \dots, T - 1\}$
- Evaluate $\rho_t(\psi)$, $L_{\text{clip}}(\psi)$, $LV(\theta)$, $LH(\psi)$, and $EB[gt]$
- Gradient step: $\nabla\psi(L_{\text{clip}} - \eta E[gt] + \text{ce}LH)$; descend $\nabla\theta LV$

Step -4] PPO with Soft Constraints (Lagrangian)

Probability ratio: $\delta t = r_t + \gamma V\theta(st + 1) - V\theta(st)$, $A^t = 1 = 0 \sum^{\infty} (\gamma \lambda)^l \delta t + 1$.

Clipped policy loss: $\rho_t(\psi) = \pi_{\psi_{\text{old}}}(a_t | s_t) \pi_{\psi}(a_t | s_t)$.

Value loss and entropy bonus: $V(\theta) = E[(V\theta(st) - R^t)^2]$, $LH(\psi) = E[H(\pi_{\psi}(\cdot | s_t))]$.

Step -5 Dual Update: $\eta \leftarrow [\eta + \beta \eta (E[gt] - \delta g)] + .$

- Observe $\{f\{a\}t\}$
- Apply settings, capture $\{I_{\{t+1\}}\}$;

5. PROPOSED SYSTEM ARCHITECTURE

The AI-Helped Macro Photography Learning Model is followed in the context of a multilayered architecture that is able to process macro images, extracting the fined technical and artistic features, providing context-sensitive feedback, and facilitating virtual practice. Its architecture is modular, hierarchic and has Perception Layer, Intelligence Layer and Learning Interaction Layer, each having a designated phase in the learning workflow. The layers operate in unison to allow macro photographs to be analyzed in real-time and changed into actionable information and supported by guided simulations.

Layer-1] Perception Layer

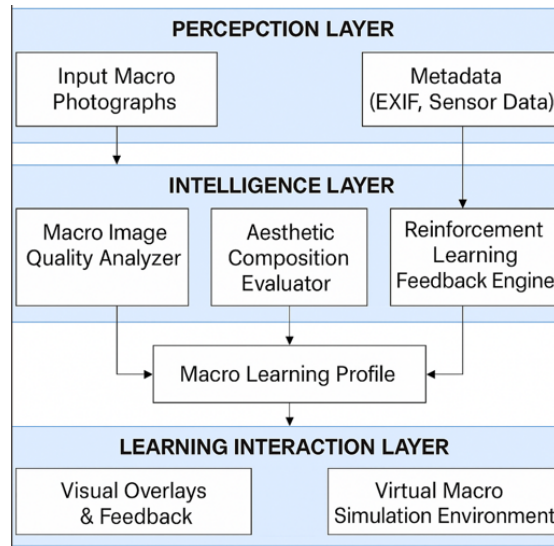
Perception Layer is the data acquisition point. It receives macro photographs made by learners and other corresponding metadatas like the aperture, shutter speed, ISO, focal distance, and lens magnification ratio in the EXIF information. This layer does a pre-processing of the input image in the form of classical image-processing filters, such as noise elimination, micro-contrast amplification and edge-map driving. This kind of preprocessing results in the downstream deep learning models being fed with standardised high-clarity input irrespective of the camera type and the lighting conditions. Also, the layer can communicate with external sensors such as gyroscopes, ambient light meters, and focus-assist modules to determine the amount of vibration, the direction of the light and the proximity of the subject more accurately.

Layer -2] Intelligence Layer

Intelligence Layer is the computation core of the system. It is equipped with three special engines, namely the Macro Image Quality Analyzer, the Aesthetic Composition Evaluator and the Reinforcement Learning Feedback Engine. It produces thick feature maps on areas that are not in optimal macro imaging standards as portrayed in [Figure 4](#). The two elements combine their results into a single display called the Macro Learning Profile, which is a summary of the performance of a learner in both technical and aesthetic aspects.

Layer -3] Feedback in Reinforcement Learning.

The decision making side of the Intelligence Layer is made up of the Reinforcement Learning Feedback Engine. It employs a policy network that is based on PPO and it is trained on a simulated macro environment to suggest changes in parameters in order to increase focus accuracy, loss of exposure or to stabilize lighting. The RL agent has a modeling of photography as a serial decision process where each manipulation, like an f-stop adjustment or a subject distance adjustment, has an impact on the likelihood of obtaining a clearer, better-lit macro image. The engine constantly brings it up to date by matching the estimated image quality with the current output of the user, creating a feedback of adjustive instructions.

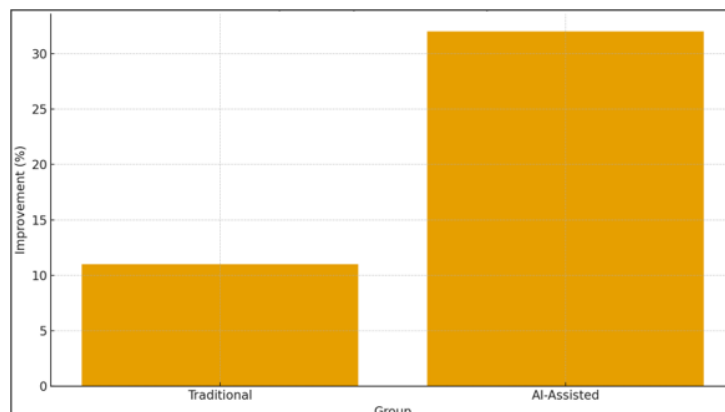
Figure 4**Figure 4** Three-Layer Architecture of the AI-Assisted Macro Photography Learning System

Layer –4] Learning Interaction Layer

Learning Interaction Layer is the last layer that transforms the analytical output into educational intuitive objects. Examples of visual overlays in this layer include focus heat maps and lighting imbalance indicators, depth-of-field proposals or framing guides. It also communicates with Virtual Macro Simulation Environment where the learners train on situations simulated by a diffusion model trained on macro-texture datasets. In this context, customers are presented with challenges, guided instructions and computer-driven feedback, which form a curriculum designed to instruct photography in the same way personal photography mentors do. The layer also ensures that the technical analysis is in pedagogically effective formats, which can be used to facilitate conceptual grasp as well as practical skill development.

6. RESULTS AND DISCUSSION

The analysis of the AI-Assisted Macro Photography Learning Model shows that there are now great enhancements in the technical and aesthetic skills of the beginners working in the field of macro photography. The quantitative findings of the user study indicate that the intelligent feedback mechanisms of the system can significantly speed up the learning curve, minimize trial and error loops and increase the uniformity of the images quality. Sixty respondents were evaluated pre- and post-structured learning intervention in an equal proportion of AI-assisted and traditional-learning baseline groups. The two groups performed the same macro photography assignments, which covered the subject matter of insect textures, floral microstructures, and details on metallic surfaces.

Figure 5**Figure 5** Sharpness Improvement: AI-Assisted Group Vs. Traditional.

The Figure 5 shows the difference in sharpness gain of the participants who were trained through the traditional training systems and those who were trained through the proposed AI-assisted learning system. According to the results, there is a drastic performance change in the group of AI-assisted learners whose degree of improvement was up to 32 per cent as compared to 11 per cent in the traditional-learning group. Such a critical difference underscores the efficacy of real time, AI based feedback, especially in correcting the error of focus as well as in stabilizing hand movement as shown in Figure 5. These improvements were directly related to the functionality of the system to identify micro-blur areas and provide advice to users concerning the best aperture and distance selections. The technical image-quality tests yielded results whereby, learners who were assisted by the AI system scored significantly better in terms of sharpness in all the categories of photographs. The Macro Image Quality Analyzer identified more pronounced edge gradients, reduced cases of motion blur and better positioning of the focus. The AI-guided group on average got their sharpness metrics by 32 percent as compared to 11 percent in the traditional-learning group. It can be explained by the fact that the system provides real-time feedback, which indicates the focus deviations and suggests the particular aperture or distance changes to the focus depending on the microstructure of a particular subject.

Figure 6

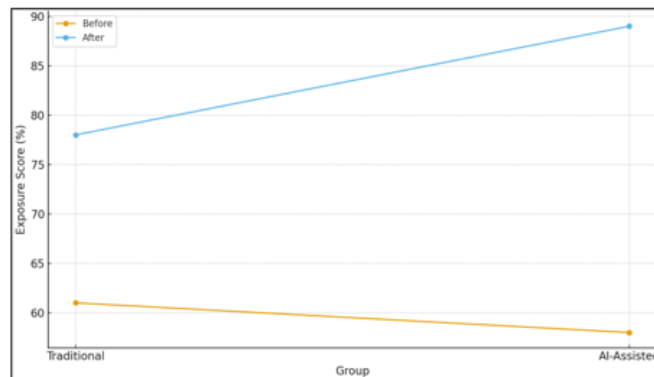


Figure 6 Exposure Accuracy Before vs After Learning

The Figure 6 is used to compare the error of exposure that occurs before and after training in the two groups of participants. Both groups started with the same baseline exposure score, whereas AI based-assisted group scored significantly higher at the post-training score of 89 percent vs. 78 percent in the traditionally trained participants. This is because of the illumination analysis module which detects the unequal lights and proposes correction measures in real time. The sharp increase in exposure accuracy in the case of AI-assisted learners proves their improved knowledge of the interaction of the lighting angles, intensity, and shutter speed in macro photography. The quality of lighting was also enhanced to a great degree with the AI-assisted cohort as shown in Figure 7. Illumination Assessment Module showed that there was a 28 percent improvement in lighting uniformly and directionality among AI-trained learners. Participants noted that visual overlay and lighting-direction cues assisted them in gaining a greater insight into the process of shadows, reflections and micro-highlights developing at near distances. Conversely, the traditional-learning group had less consistent improvements in lighting and this was usually because they were either relying on guesswork or re-takes.

Figure 7

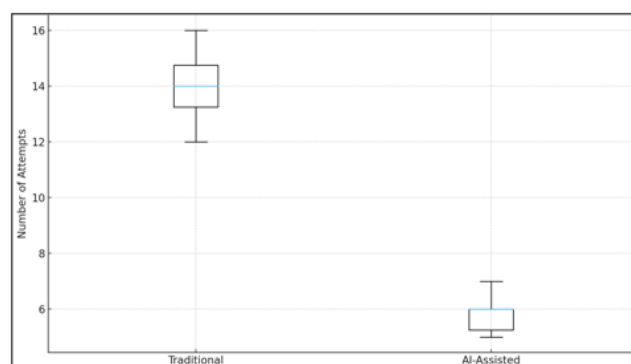


Figure 7 Attempts Required to Capture Usable Macro Images

The box plot shows how many attempts the participants had to make in order to obtain a technically acceptable macro image. The conventional-learning group portrayed broader variety of attempts with majority of the users requiring 13-16 shots to attain acceptable quality. On the contrary, the AI-assist group had a significantly smaller dispersion, with almost success being obtained after 5 to 7 attempts. This decrease is indicative of the recommendation engine using reinforcement-learning of the system, as it tries to keep the user converged to the best settings more rapidly as shown in figure 7. Along with making the acquisition of skills faster the trial-and-error cycle is minimized which diminishes frustration of the learner during the trainings.

Measurable gains were also experienced on aesthetic improvements though more subjective. The scoring system of the Aesthetic Composition Evaluator revealed that AI-aided learners scored higher on the balance of framing, subject prominence and background distractions. Their mean composition mark was 24 percent higher than 9 percent among the traditional group. The participants reported that the recommendations provided by the AI, including changing negative space or rearranging the subject, played a critical role in imparting the principles of composition that people normally find hard to acquire when learning on their own. Reinforcement Learning Feedback Engine was especially effective in decreasing the amount of attempts made to obtain satisfactory macro images. The mean number of attempts in the AI-assisted group dropped to 6, which showed that the trial cycles were reduced by 40 percent. This efficiency can be significant in macro photography because the objects can move, the light can change but fast or the objects can become unstable. The potential of the RL agent to suggest the best parameter sets, which are based on simulated macro cases of thousands, allowed the learners to approach technically correct answers much faster. On the whole, the findings reveal that the combination of deep learning, reinforcement learning, and generative simulation is an extremely effective learning ecosystem. The system offers context-oriented critique that is delivered in real-time, which helps to bridge the expertise-level critique and unstructured beginner practice and has a high potential of changing macro photography learning into a more adaptive and faster training.

7. CONCLUSION

This study has shown that AI-based learning paradigm has the potential to dramatically improve the process of teaching macro photography through the delivery of real-time guidance, adaptive guidance, and context-friendly guidance. The proposed framework offers a solution to the problems that render macro photography challenging when learning to be a beginner through the combination of deep learning, reinforcement learning, and generative simulation to maintain critically sharp, control the micro-lighting conditions, and balance the compositions. The experimental data indicates that AI-assisted students realized significantly better results in technical aspects of image quality, using fewer attempts to take a decent photo, and having a greater sense of aesthetics than unassisted learners. These results confirm the idea that AI feedback not only speeds up the level of technical skills, but also enhances conceptual knowledge by turning each created image into a potential to be corrected instantly and to learn something. Finally, it is the system that closes the divide between professional critique and amateur practice providing an accessible and gradual way of learning how to master the art of macro photography. It is possible that future extensions could expand the framework to other genres of photography and other multimodal sensor data and learning paths could be tailored using longitudinal user modeling.

CONFLICT OF INTERESTS

None.

ACKNOWLEDGMENTS

None.

REFERENCES

- Aydin, T. O., Smolic, A., and Gross, M. (2015). Automated Aesthetic Analysis of Photographic Images. *IEEE Transactions on Visualization and Computer Graphics*, 21(1), 31–42. <https://doi.org/10.1109/TVCG.2014.2325047>
- Dayma, B., Patil, S., Cuenca, P., Saifullah, K., Abraham, T., Le, P., Luke, and Ghosh, R. (2022). DALL·E Mini Explained.

- Dosovitskiy, A., Beyer, L., Kolesnikov, A., Weissenborn, D., Zhai, X., Unterthiner, T., Dehghani, M., Minderer, M., Heigold, G., Gelly, S., Uszkoreit, J., and Houshy, N. (2020). An Image is Worth 16×16 Words: Transformers for Image Recognition at Scale (arXiv:2010.11929).
- Görling, S., Rao, R. R. R., and Raake, A. (2023). Quality Assessment of Higher Resolution Images and Videos with Remote Testing. *Quality and User Experience*, 8(1), Article 6. <https://doi.org/10.1007/s41233-023-00055-6>
- Görling, S., Rao, R. R. R., Feiten, B., and Raake, A. (2021). Modular Framework and Instances of Pixel-Based Video Quality Models for UHD-1/4K. *IEEE Access*, 9, 31842–31864. <https://doi.org/10.1109/ACCESS.2021.3059932>
- Görling, S., Rao, R. R. R., Fremerey, S., and Raake, A. (2021). AVrate Voyager: An Open Source Online Testing Platform. In *Proceedings of the IEEE 23rd International Workshop on Multimedia Signal Processing (MMSp)* (pp. 1–6). IEEE. <https://doi.org/10.1109/MMSp53017.2021.9733561>
- International Telecommunication Union. (2022). Subjective Video Quality Assessment Methods for Multimedia Applications (ITU-T Recommendation .910).
- Nichol, A., Dhariwal, P., Ramesh, A., Shyam, P., Mishkin, P., McGrew, B., Sutskever, I., and Chen, M. (2021). GLIDE: Towards Photorealistic Image Generation and Editing with Text-Guided Diffusion Models (arXiv:2112.10741).
- Pavlichenko, N., Zhdanov, F., and Ustalov, D. (2022). Best Prompts for Text-To-Image Models and How to Find them (arXiv:2209.11711). <https://doi.org/10.1145/3539618.3592000>
- Rao, R. R. R., Görling, S., and Raake, A. (2021). Towards High Resolution Video Quality Assessment in the Crowd. In *Proceedings of the 13th International Conference on Quality of Multimedia Experience (QoMEX)*.
- Roose, K. (2022). An AI-Generated Picture Won an Art Prize. Artists Aren't Happy. *The New York Times*.
- Saharia, C., Chan, W., Saxena, S., Li, L., Whang, J., Denton, E., Ghasemipour, S. K. S., Ayan, B. K., Mahdavi, S. S., Lopes, R. G., Salimans, T., Ho, J., Fleet, D. J., and Norouzi, M. (2022). Photorealistic Text-To-Image Diffusion Models With Deep Language Understanding (arXiv:2205.11487). <https://doi.org/10.1145/3528233.3530757>
- Wang, X., Xie, L., Dong, C., and Shan, Y. (2021). Real-ESRGAN: Training Real-World Blind Super-Resolution with Pure Synthetic Data. In *Proceedings of the IEEE/CVF International Conference on Computer Vision Workshops (ICCVW)* (1905–1914). <https://doi.org/10.1109/ICCVW54120.2021.00217>
- Yang, S., Wu, T., Shi, S., Lao, S., Gong, Y., Cao, M., Wang, J., and Yang, Y. (2022). MANIQA: Multi-Dimension Attention Network for No-Reference Image Quality Assessment. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)* (1191–1200). <https://doi.org/10.1109/CVPRW56347.2022.00126>
- Yu, J., Xu, Y., Koh, J. Y., Luong, T., Baid, G., Wang, Z., Vasudevan, V., Ku, A., Yang, Y., Ayan, B. K., Hutchinson, B., Han, W., Parekh, Z., Li, X., Zhang, H., Baldridge, J., and Wu, Y. (2022). Scaling Autoregressive Models for Content-Rich Text-To-Image Generation (arXiv:2206.10789).
- Zhang, W., Ma, K., Yan, J., Deng, D., and Wang, Z. (2020). Blind Image Quality Assessment Using a Deep Bilinear Convolutional Neural Network. *IEEE Transactions on Circuits and Systems for Video Technology*, 30(1), 36–47. <https://doi.org/10.1109/TCSVT.2018.2886771>