Mood-Driven Colorization of Virtual Indoor Scenes Supplementary Document

Abstract—This is the supplementary document that our main paper refers to. It contains additional details of our formulation and user study results.

1 Additional User Study Data

Fig. 1: A screenshot of the user study for User Study 3. The interface is similar for User Study 1. Participants used an Oculus Rift VR headset with controllers in both User Study 1 and 3.

For a visualization of what users did during User Study 1 and 3, refer to Fig. 1. Table 3 contains the statistical results of applying the chi-squared test to the results of each scene, corresponding to the *Task* 1 and *Task* 3 user study described in Section 6 of our paper. Table 5 shows the results of all post-hoc pairwise t-tests for *Task* 2 of our User Study. For images of all twelve scenes used in User Study 1, refer to Fig. 2.

For results for each mood for one scene (Living Room), refer to Fig. 5.

For a screenshot of the User Study 2 interface, refer to Fig. 6.

For statistical information for Tasks 1 and 3 from User Study 1, please refer to Table 1.

For the distribution of votes from Task 2 from User Study 1, refer to Table 2.

Fig. 7 to Fig. 11 contain a sample of scenes that were used to train our classifier organized by mood. We provide links where these images can be viewed.

For all scenes and results related to the MTurk user study, refer to Fig. 12 and Fig. 13.

2 Explanation of Color Distance Metric mentioned in Section 5.3 of Main Paper

We define metric $D(.)$ as follows:

$$D(\phi^c, \psi^c) = \frac{1}{100}\sum_{\mathbf{u}\in\phi^c} \min\{w_{\mathbf{u}} \cdot d(\mathbf{u},\mathbf{v}) \mid \mathbf{v}\in\psi^c\}, \tag{1}$$

where $\mathbf{u}$ and $\mathbf{v} \in \mathbb{R}^3$ are the dominant colors in ϕ^c and ψ^c respectively, $w_{\mathbf{u}}$ is the weight which is the proportion of the color $\mathbf{u}$ in the five-dominant-colors ϕ^c, and $d(.)$ is the metric to determine if the two colors are distinguishable to the human eye.

$$d(u,v) = \sqrt{(\frac{\Delta L'}{S_L})^2 + (\frac{\Delta C'}{S_C})^2 + (\frac{\Delta H'}{S_h})^2 + R_T \frac{\Delta C'}{S_C}\frac{\Delta H'}{S_h}}, \tag{2}$$

where $\Delta L' = v_{L^*} - u_{L^*}$. $\Delta C' = v_{C'} - u_{C'}$ where $u_{C'} = \sqrt{u_{a'^2} + u_{b^{*2}}}$ and $v_{C'}$ is defined analogously. $u_{b^{*2}}$ and $v_{b^{*2}}$ refer to the b^* component of each color in CIELab space. $u_{a'}$ is defined as follows:

$$u_{a'} = u_{a^*} + \frac{u_{a^*}}{2}(1 - \sqrt{\frac{\tilde{C}^7}{\tilde{C}^7 + 25^7}}), \tag{3}$$

where a_2' is defined analogously. $\tilde{C} = \frac{u_{C^*} + v_{C^*}}{2}$ here. The component $\Delta H' = 2\sqrt{u_{C'} v_{C'}} \sin(\frac{\Delta h'}{2})$. $\Delta h'$ is defined as follows:

$$\Delta h' = \begin{cases} v_{h'} - u_{h'}, & \text{if } |u_{h'} - v_{h'}| \le 180° \\ v_{h'} - u_{h'} + 360°, & \text{if } |u_{h'} - v_{h'}| > 180° \text{ and } v_{h'} \le u_{h'} \\ v_{h'} - u_{h'} - 360°, & \text{if } |u_{h'} - v_{h'}| > 180° \text{ and } v_{h'} > u_{h'}. \end{cases}$$

Here $u_{h'} = \arctan 2(u_{b^*}, u_{a'}) \mod 360$ and $v_{h'}$ is defined similarly. $S_L = 1 + \frac{0.015(\tilde{L}-50)^2}{\sqrt{20+(\tilde{L}-50)^2}}$, $\tilde{L} = \frac{u_{L^*} + v_{L^*}}{2}$, $S_c = 1 + 0.045\tilde{C}'$, $\tilde{C}' = \frac{u_{C'} + v_{C'}}{2}$ and $S_H = 1 + 0.015\tilde{C}'T$. T is defined as follows:

$$\begin{aligned} T = 1 &- 0.17\cos(\tilde{H}' - 30°) + 0.24\cos(2\tilde{H}') \\ &+ 0.32\cos(3\tilde{H}' + 6°) - 0.20\cos(4\tilde{H}' - 63°), \end{aligned} \tag{4}$$

and $\tilde{H}'$ is defined as follows:

$$\tilde{H}' = \begin{cases} \frac{v_{h'} + u_{h'}}{2}, & \text{if } |u_{h'} - v_{h'}| \le 180° \\ \frac{v_{h'} + u_{h'} + 360°}{2}, & \text{if } |u_{h'} - v_{h'}| > 180° \text{ and } v_{h'} + u_{h'} < 360° \\ \frac{v_{h'} + u_{h'} - 360°}{2}, & \text{if } |u_{h'} - v_{h'}| > 180° \text{ and } v_{h'} + u_{h'} \ge 360°. \end{cases}$$

Finally, the hue rotation term R_t is defined as follows:

$$R_T = -2\sqrt{\frac{\tilde{C}^7}{\tilde{C}^7 + 25^7}} \sin(60° * \exp(-(\frac{\tilde{H}' - 275°}{25°})^2)). \tag{5}$$

We have described the general CIEDE2000 algorithm as applied in our work. The range of $d(.)$ is $[0,100]$ and the sum of $w_{\mathbf{u}}$, where $\mathbf{u} \in \phi^c$, is 1, so the range of Equation 1 is $[0,100]$, hence a normalization 100. The lower the output of $d(.)$ the closer the colors match.

3 User Study 4 Additional Information

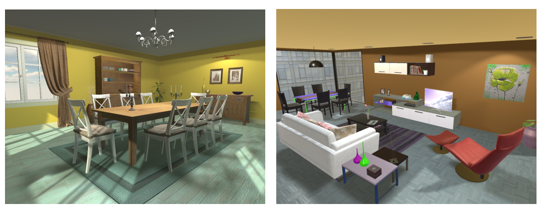
Fig 14 shows a screenshot of the User Study 4. The distribution of votes for these questions asked during the user study can be seen in Fig 15.

(a) Cheerful Bedroom (b) Melancholy Bedroom (c) Peaceful Bedroom (d) Romantic Bedroom (e) Scary Bedroom

Fig. 2: Results used in *Task* 1 of User Study 1. Results generated using a mood weight of 0.8 and a realness weight of 0.2.

(a) Cheerful Dining Room (b) Scary Dining Room (c) Melancholy Living Room (d) Peaceful Bathroom (e) Romantic Living Room

Fig. 3: Results used in *Task* 2 of User Study 1. Results generated using a mood weight of 0.8 and a realness weight of 0.2.



(a) Realistic Dining Room (b) Realistic Living Room

Fig. 4: Results used in *Task* 3 of User Study 1. Here results were generated using a mood weight of 0.0 and a realness weight of 1.0.

(a) Cheerful Living Room (b) Melancholy Living Room (c) Peaceful Living Room (d) Romantic Living Room (e) Scary Living Room

Fig. 5: Optimization results for the Dining Room scene. Results generated using a mood weight of 0.8 and a realness weight of 0.2.

Fig. 6: A screenshot of the task that Turkers had to complete in our user study. In this example, the result generated by our approach is on the left and the result created by an artist is on the right. Participants completed this same task for all 20 compared results. Please refer to our supplementary document for all the results that were compared.

Link	Link	Link	Link	Link
(a)	(b)	(c)	(d)	(e)
Link	Link	Link	Link	Link
(f)	(g)	(h)	(i)	(j)

Fig. 7: Examples of cheerful images from our image dataset. Click on the link to view the image.

Link	Link	Link	Link	Link
(a)	(b)	(c)	(d)	(e)
Link	Link	Link	Link	Link
(f)	(g)	(h)	(i)	(j)

Fig. 8: Examples of melancholy images from our image dataset. Click on the link to view the image.

Link	Link	Link	Link	Link
(a)	(b)	(c)	(d)	(e)
Link	Link	Link	Link	Link
(f)	(g)	(h)	(i)	(j)

Fig. 9: Examples of peaceful images from our image dataset. Click on the link to view the image.

Link	Link	Link	Link	Link
(a)	(b)	(c)	(d)	(e)
Link	Link	Link	Link	Link
(f)	(g)	(h)	(i)	(j)

Fig. 10: Examples of romantic images from our image dataset. Click on the link to view the image.

Link	Link	Link	Link	Link
(a)	(b)	(c)	(d)	(e)
Link	Link	Link	Link	Link
(f)	(g)	(h)	(i)	(j)

Fig. 11: Examples of scary images from our image dataset. Click on the link to view the image.

(a) Cheerful Bedroom 1 (23 vs. 12)

(b) Cheerful Bedroom 2 (6 vs. 29)

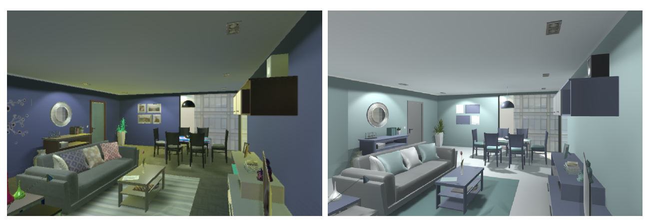
(c) Cheerful Dining Room 1 (19 vs. 16)

(d) Cheerful Dining Room 2 (5 vs. 30)

(e) Melancholy Bedroom 1 (13 vs. 22)

(f) Melancholy Bedroom 2 (15 vs. 20)

(g) Melancholy Living Room 1 (18 vs. 17)



(h) Melancholy Living Room 2 (17 vs. 18)

(i) Peaceful Bathroom 1 (17 vs. 18)

(j) Peaceful Bathroom 2 (14 vs. 21)

Fig. 12: Comparisons shown in User Study 3. Results in each pair are as follows: left: our results and votes for our result - right: same for artist results.

(a) Peaceful Bedroom 1 (17 vs. 18)

(b) Peaceful Bedroom 2 (18 vs. 17)

(c) Romantic Bedroom 1 (20 vs. 15)

(d) Romantic Bedroom 2 (9 vs. 26)

(e) Romantic Living Room (28 vs. 7)

(f) Romantic Living Room 2 (31 vs. 4)

(g) Scary Bedroom 1 (7 vs. 28)

(h) Scary Bedroom 2 (17 vs. 18)

(i) Scary Dining Room 1 (6 vs. 29)

(j) Scary Dining Room 2 (6 vs. 29)

Fig. 13: Comparisons shown in User Study 3. Results in each pair are as follows: left: our results and votes for our result - right: same for artist results.

Table 1: Statistical data on the ratings given to our synthesized results for *Task* 1 and *Task* 3 of User Study 1. The ratings refer to how convincing the participants found the results for the target mood. Ratings were given using a $5-1$ Likert scale.

Scene	Realistic Dining Room	Realistic Living Room	Cheerful Bedroom	Scary Bedroom	Melancholy Bedroom	Peaceful Bedroom	Romantic Bedroom
Mean	4.000	3.900	4.125	4.125	4.025	3.775	3.475
Mode	4.000	4.000	4.000	5.000	4.000	4.000	3.000
Standard Deviation	0.922	0.917	0.781	0.871	0.821	0.908	1.140

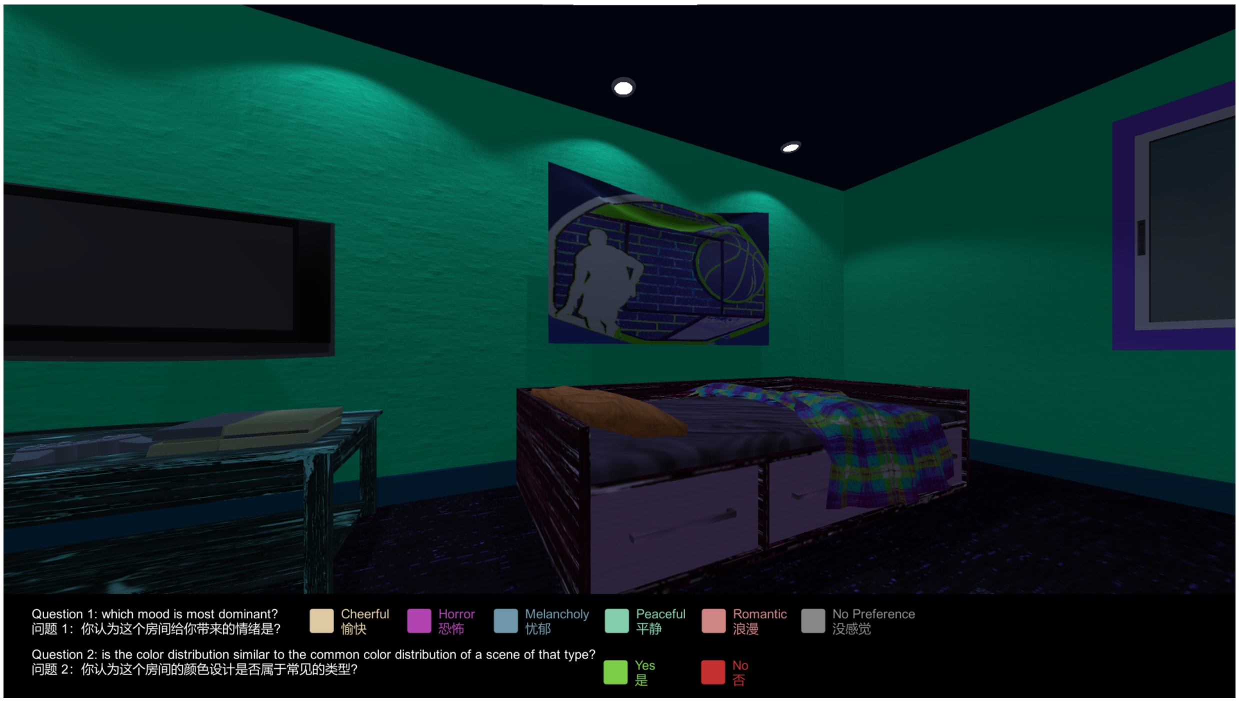


Fig. 14: A screenshot of the User Study 4.

Table 2: The number of votes for each of the 5 categorical task scenes optimized with a mood used in *Task 2* of User Study 1. The numbers in bold show the number of votes for the mood that was optimized for.

Voted Mood	Cheerful Dining Room	Scary Dining Room	Melancholy Living Room	Peaceful Bathroom	Romantic Living Room
Cheerful	**36**	0	1	14	2
Scary	1	**11**	4	0	3
Melancholy	0	19	**24**	0	0
Peaceful	1	8	11	**20**	0
Romantic	0	1	0	6	**30**
Other	2	1	0	0	5

4 CNN Features Vs. Five-Dominate Colors Features

For training the Five-Dominate Colors features classifiers, we first created a dataset comprising 5,000 images with the desired mood and 1,250 images for each mutually exclusive moods. For example, when training the cheerful classifier, we have 5,000 cheerful images and 1,250 images for each of the remaining 4 moods. We extracted the five-dominate colors from the image and fed them to a SVM classifier with an linear kernel. We trained the five classifiers for each mood in the same manner.

Table 4 shows the accuracy of the CNN features and Five-Dominate Colors features classifiers. Fig. 16 shows the results from running our optimization process with CNN features and Five-Dominate Colors features classifiers.

Table 3: P-values of chi-squared tests on categorical task scenes.

Scene	P-value
Cheerful Dining Room	9.93e−32
Scary Dining Room	5.36e−8
Melancholy Living Room	4.11e−13
Peaceful Bathroom	1.44e−10
Romantic Living Room	3.76e−20

Table 4: The accuracy of the CNN features and Five-Dominate Colors features classifiers for different moods

Classifier	CNN	5DC
Cheerful	**85.80%**	85.10%
Scary	**96.75%**	95.45%
Melancholy	**78.90%**	78.85%
Peaceful	84.65%	**84.75%**
Romantic	**86.20%**	81.80%

In general, for the Five-Dominate Colors features classifiers, certain moods are difficult to distinguish, such as cheerful and peaceful, as well as scary and melancholy. These moods may be similar in the color distribution. Fig 16(m) and (n) shows two examples of the confusion. If we want to improve the mood classification accuracy for the Five-Dominate Colors features classifiers, material, textures, and other visual information may be needed as additional handcrafted features.

Table 5: P-values from conducting t-tests on quantitative task scenes. Values in bold show a statistically significant difference.

Scenes	Realistic Dining Room	Realistic Living Room	Cheerful Bedroom	Scary Bedroom	Melancholy Bedroom	Peaceful Bedroom	Romantic Bedroom
Realistic Dining Room	-	0.552	0.443	0.540	0.563	0.226	**0.025**
	Realistic Living Room	-	0.163	0.223	1.000	0.545	0.081
		Cheerful Bedroom	-	0.895	0.177	**0.039**	**0.003**
			Scary Bedroom	-	0.237	0.063	**0.005**
				Melancholy Bedroom	-	0.556	0.0880
					Peaceful Bedroom	-	0.209
						Romantic Bedroom	-

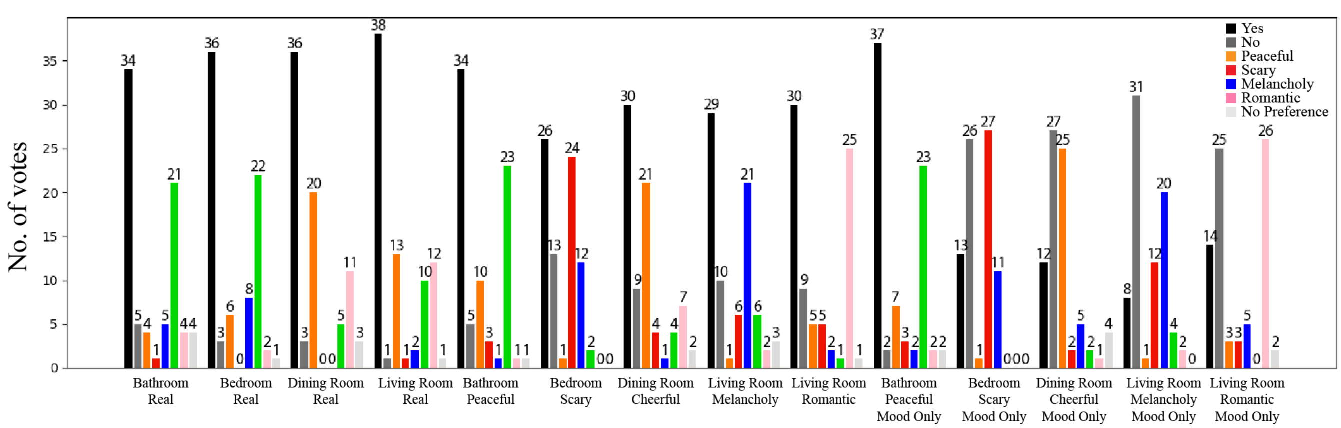


Fig. 15: Results of the User Study 4

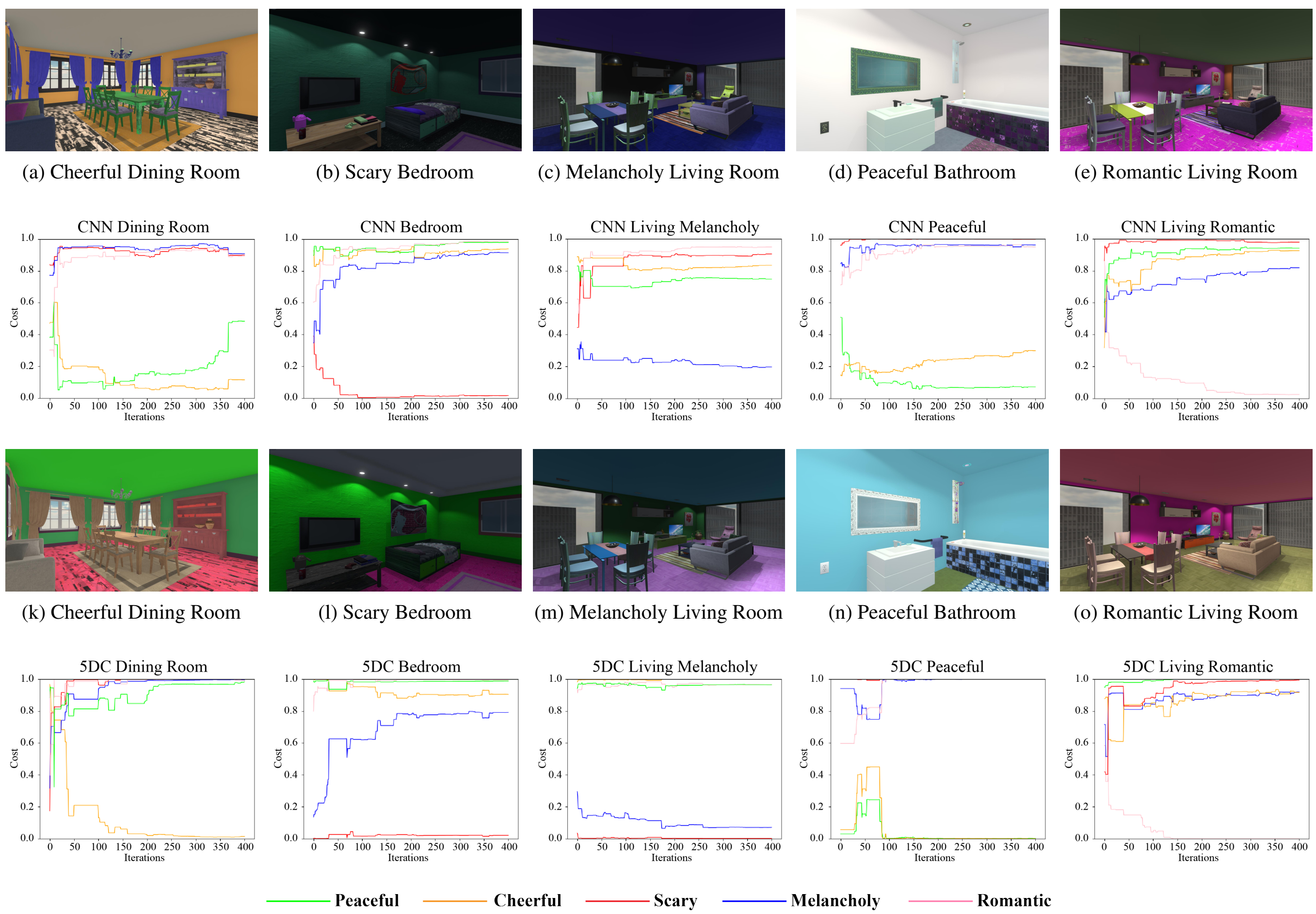


Fig. 16: (a-e) Results synthesized by running our optimization process with the CNN features classifiers. (k-o) Results synthesized by running our optimization process with the Five-Dominate Colors features classifiers. The plot (m) of "Melancholy Living Room" generated using the Five-Dominate Colors classifier shows that the Five-Dominate Colors melancholy classifier cannot distinguish the scary mood colors and the melancholy mood colors well; therefore both scary mood cost and melancholy mood cost drops to a low level while only melancholy was used as the target mood. Similarly, the plot (n) of the "Peaceful Bathroom" shows that the Five-Dominate Colors peaceful classifier cannot distinguish peaceful mood colors and the cheerful mood colors well.