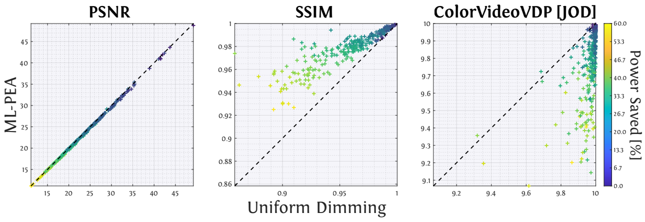

Figure 8: *The x-axis represents metric scores for uniform dimming, and the y-axis metric scores for ML-PEA. Scatter color is power saved (%), and the identity line is plotted as a black dashed line. Higher scores equal higher predicted quality.*

8. Image Quality Metric Results

We ran three metrics on the test dataset from DIV2k: PSNR, SSIM, and ColorVideoVDP. The results are shown in Figure 8, where each scatter point corresponds to one of the 100 test images from the DIV2k test dataset. Note that this plot shows results for the three target power saving rates, so in total there are 300 scatter points for each metric. A model can be considered to perform well if scatter points lie above the identity line, i.e. metric scores for ML-PEA output is higher than scores for uniformly-dimming images.

Selection of PSNR and SSIM as evaluation metrics is based on prior art, which use these metrics to evaluate their models [SDLM24, ADMM25, LMDB23]. The inclusion of ColorVideoVDP [MHA*24] was meant to introduce a modern metric based on low-level models of human vision, trained on display-related distortions.

9. Additional Ablation Studies

Here, we discuss additional ablation studies and experiments, meant to supplement the discussion from Section 4.2.

9.1. Ablating Loss Weights

We conducted an ablation on the loss function weights, as described in Section 4.2. The following weights were studied in this experiment:

- $\lambda_{\mathcal{P}}$: $\{5.0,\ 50.0\}$
- λ_{VGG} : $\{0.0,\ 0.05,\ 0.5\}$
- λ_{SSIM} : $\{0.0,\ 0.5,\ 5.0\}$.

The results for each combination of parameters are shown, for power saving targets of 17%, 32%, and 45% ($\alpha = \{0.83,\ 0.68,\ 0.45\}$), in Table 2. Cell colors represent first, second, and third-best performance. Note that the column "Power Target - Pred." represents the quantity

$$\Delta\mathcal{P} = 100 \cdot (1-\alpha) - 100 \cdot (1 - \mathcal{P}(\mathcal{I}^*)/\mathcal{P}(\mathcal{I})), \tag{9}$$

which is essentially the difference in power savings between the optimized image $\mathcal{I}^*$ and the target power saving rate. It is important that the model outputs images which closely match the target power saving rate, and so models that have a value of $\Delta\mathcal{P}$ close to 0 are ideal. Here, we recall that α is the target proportion of power consumed by the target, relative to the input. In other words, if we define T as the target power savings (%), then $T = 100 \cdot (1-\alpha)$.

One important note is that the rankings in Table 2 do not paint a complete story – while we show which combination of parameters perform best in terms of a number of common image quality metrics, these naturally depend on the accuracy of the model to produce images with power savings close to the target. In other words, when inspecting Table 2 we notice that PSNR, SSIM, and CVVDP scores are typically highest for $\Delta\mathcal{P}$ with large magnitude (or models that do not approximate the target power savings well). As a result, it is important to jointly consider $\Delta\mathcal{P}$ as well as the metric scores to find a fine balance between the two when selecting model parameters. The ability to control the power savings of the model's output is crucial to its performance, and the core problem in our constrained optimization. In our experiments, we used the parameters of the last row in each α block ($\lambda_{\mathcal{P}} = 50.0$, $\lambda_{\text{VGG}} = 0.5$, and $\lambda_{\text{SSIM}} = 5.0$).

We make the decision to display Table 2 with ML-PEA and uniform dimming results side by side and mark the rankings within techniques, rather than between techniques. The reason for this is because we want to show optimal parameters for ML-PEA. Comparisons between ML-PEA and uniform dimming can still be made by comparing the results within the same row.

9.2. Ablating Element-Wise Dimming Map Application

In Section 3, we allude to the fact that our element-wise multiplication (MULT) function f is optimal compared to the addition (ADD) operation used in prior art. We conducted an ablation on f (ADD or MULT) as well as the number of channels (1 or 3) in the output dimming map. The results of these experiments are shown in Table 3. A qualitative comparison is shown in Figure 9, where we can

Table 2: *Here, we display results of ablations on the weights of loss functions used in our experiments. We show results for both ML-PEA (●) and uniform dimming (●). Cell colors represent first, second, and third-best performance in each α block within power saving techniques.*

α	$\lambda_{\mathcal{P}}$	λ_{VGG}	λ_{SSIM}	$\Delta\mathcal{P}$	PSNR ↑ (●)	SSIM ↑ (●)	CVVDP ↑ (●)	PSNR ↑ (●)	SSIM ↑ (●)	CVVDP ↑ (●)
0.55	5.00	0.00	0.50	1.287	18.948	0.966	8.928	19.380	0.932	9.928
	50.0	0.00	0.50	2.495	19.398	0.967	9.050	19.551	0.939	9.937
	5.00	0.00	5.00	16.506	23.271	0.987	9.374	23.759	0.975	9.979
	50.0	0.00	5.00	0.855	18.838	0.964	8.843	19.273	0.931	9.927
	5.00	0.05	0.00	2.816	19.495	0.951	9.755	19.765	0.939	9.940
	50.0	0.05	0.00	-1.617	18.510	0.926	9.836	18.486	0.924	9.913
	5.00	0.05	0.50	4.764	19.963	0.972	9.558	20.208	0.945	9.943
	50.0	0.05	0.50	-3.785	17.944	0.955	9.467	18.123	0.912	9.904
	5.00	0.05	5.00	14.375	22.681	0.985	9.694	23.008	0.971	9.974
	50.0	0.05	5.00	3.634	19.662	0.971	9.405	19.877	0.942	9.940
	5.00	0.50	0.00	-3.914	17.763	0.940	9.228	18.464	0.903	9.893
	50.0	0.50	0.00	-6.029	17.403	0.921	9.440	17.761	0.898	9.888
	5.00	0.50	0.50	6.016	20.276	0.971	9.469	20.968	0.945	9.943
	50.0	0.50	0.50	-9.064	16.767	0.936	9.356	17.035	0.882	9.862
	5.00	0.50	5.00	15.778	23.192	0.987	9.846	23.442	0.974	9.979
	50.0	0.50	5.00	0.092	18.805	0.963	9.551	19.076	0.928	9.924
0.68	5.00	0.00	0.50	2.044	22.829	0.986	9.352	23.170	0.972	9.976
	50.0	0.00	0.50	0.886	22.454	0.981	9.098	22.602	0.971	9.975
	5.00	0.00	5.00	9.554	25.574	0.992	9.440	25.900	0.985	9.989
	50.0	0.00	5.00	-0.038	22.131	0.983	9.167	22.553	0.968	9.970
	5.00	0.05	0.00	-5.509	20.695	0.961	9.763	20.904	0.955	9.958
	50.0	0.05	0.00	-0.688	22.108	0.970	9.927	22.078	0.968	9.970
	5.00	0.05	0.50	1.390	22.729	0.985	9.786	23.011	0.971	9.975
	50.0	0.05	0.50	-2.376	21.614	0.980	9.865	21.672	0.963	9.967
	5.00	0.05	5.00	9.627	25.736	0.992	9.817	26.008	0.985	9.989
	50.0	0.05	5.00	1.080	22.621	0.985	9.736	22.838	0.971	9.974
	5.00	0.50	0.00	6.627	24.647	0.985	9.765	25.421	0.979	9.983
	50.0	0.50	0.00	-6.007	20.551	0.969	9.707	20.853	0.952	9.956
	5.00	0.50	0.50	4.437	23.778	0.987	9.855	24.171	0.977	9.982
	50.0	0.50	0.50	-3.243	21.341	0.979	9.758	21.626	0.960	9.963
	5.00	0.50	5.00	13.992	27.434	0.993	9.744	27.976	0.990	9.995
	50.0	0.50	5.00	-1.613	21.818	0.982	9.812	21.963	0.965	9.968
0.83	5.00	0.00	0.50	-0.046	28.038	0.994	9.443	28.441	0.991	9.994
	50.0	0.00	0.50	-0.413	27.961	0.993	9.389	28.014	0.991	9.994
	5.00	0.00	5.00	4.733	31.006	0.996	9.568	31.444	0.995	9.997
	50.0	0.00	5.00	-1.445	27.367	0.994	9.428	27.618	0.990	9.994
	5.00	0.05	0.00	-4.332	26.196	0.990	9.891	26.555	0.987	9.991
	50.0	0.05	0.00	-0.876	27.814	0.992	9.982	27.775	0.991	9.994
	5.00	0.05	0.50	1.609	29.257	0.996	9.942	29.535	0.993	9.995
	50.0	0.05	0.50	-0.759	27.897	0.994	9.979	27.878	0.991	9.994
	5.00	0.05	5.00	4.425	31.050	0.997	9.969	31.280	0.995	9.997
	50.0	0.05	5.00	-0.374	28.060	0.995	9.819	28.302	0.991	9.994
	5.00	0.50	0.00	-2.207	27.237	0.989	9.891	27.949	0.988	9.993
	50.0	0.50	0.00	-3.702	26.425	0.990	9.895	26.657	0.988	9.992
	5.00	0.50	0.50	1.741	29.279	0.995	9.924	29.775	0.993	9.996
	50.0	0.50	0.50	-1.392	27.631	0.994	9.928	27.937	0.990	9.994
	5.00	0.50	5.00	7.636	33.662	0.998	9.989	33.748	0.997	9.997
	50.0	0.50	5.00	-0.718	27.937	0.995	9.926	28.222	0.991	9.994

Table 3: *Ablation study results on the elementwise dimming map application, f, and the number of channels in the dimming map, C, are shown here.*

α	f	C	$\Delta\mathcal{P}$	PSNR ↑ (●)	SSIM ↑ (●)	CVVDP ↑ (●)	PSNR ↑ (●)	SSIM ↑ (●)	CVVDP ↑ (●)
0.55	ADD	3	0.537	18.566	0.955	9.074	19.166	0.930	9.933
	MULT	3	3.543	19.599	0.972	9.560	19.930	0.940	9.939
	ADD	1	-0.044	17.957	0.947	8.527	19.414	0.920	9.912
	MULT	1	5.976	20.244	0.974	9.590	20.586	0.948	9.949
0.68	ADD	3	-1.610	21.523	0.982	9.490	22.221	0.963	9.966
	MULT	3	1.199	22.642	0.985	9.835	22.856	0.971	9.975
	ADD	1	3.433	22.963	0.985	9.577	24.172	0.973	9.977
	MULT	1	-1.086	21.953	0.983	9.722	22.183	0.966	9.969
0.83	ADD	3	-1.360	27.405	0.995	9.890	27.900	0.990	9.994
	MULT	3	-0.483	27.976	0.995	9.910	28.168	0.991	9.995
	ADD	1	-10.50	23.155	0.985	9.397	24.570	0.975	9.979
	MULT	1	0.751	28.745	0.995	9.958	28.988	0.992	9.995

see visible artifacts around edge features in the 3-channel, ADD condition which are not visible in the single-channel MULT one. We find from this experiment that the MULT operator f performs best. For all metrics, the 3-channel dimming map with MULT performed 2nd-best or better for all target power saving rates. The 1-channel dimming map with MULT performed best for target power saving rates of 45% and 83%, and performed in the top 3 for a 32% savings target.

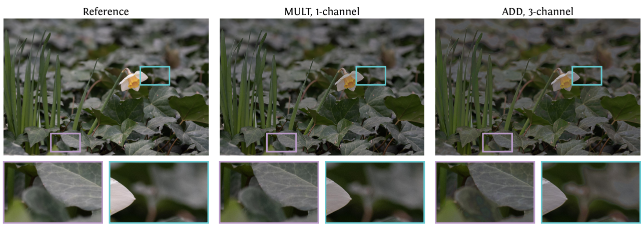

Figure 9: *We ablate the number of dimming map channels and the element-wise function f for applying the dimming map to input images.*

9.3. Comparisons with Le Meur et al. (2023)

We conducted a comparison between our ML-PEA technique and that of [LMDB23], which is the most recent and relevant prior machine learning approach to display power optimization. There was no open source code, so we attempted to replicate their pipeline as best as possible. Their technique optimizes four loss functions: an L1, SSIM, and power loss between the input and output images, as well as a total variation loss on the dimming map. They also used an ADD operation to apply output 1-channel dimming maps to the input images.

We found that the metric scores of [LMDB23] were lower compared to uniform dimming and ML-PEA for the three target power saving rates we studied, as shown in Table 4. In addition, we make a plot for this table, shown in Figure 10, to visualize the result. We do this because the power saving rates are not matched between ML-PEA and [LMDB23]. It is clear that the method of [LMDB23] performs worse in terms of the three image quality metrics.

Table 4: *Average scores are tabulated for uniform dimming, ML-PEA, and [LMDB23].*

α	Method	PSNR (dB) ↑	SSIM ↑	CVVDP (JOD) ↑	Power Saved
0.55	Uniform Dimming	19.08	0.93	9.92	-
	ML-PEA	18.81	0.96	9.55	44.91%
	Uniform Dimming	17.46	0.89	9.83	-
	[LMDB23]	16.02	0.67	8.11	51.76%
0.68	Uniform Dimming	21.96	0.96	9.97	-
	ML-PEA	21.82	0.98	9.81	33.61%
	Uniform Dimming	21.02	0.95	9.93	-
	[LMDB23]	19.39	0.78	8.89	36.98%
0.83	Uniform Dimming	28.22	0.99	9.99	-
	ML-PEA	27.94	0.99	9.93	17.72%
	Uniform Dimming	26.29	0.99	9.98	-
	[LMDB23]	24.79	0.90	9.63	21.26%

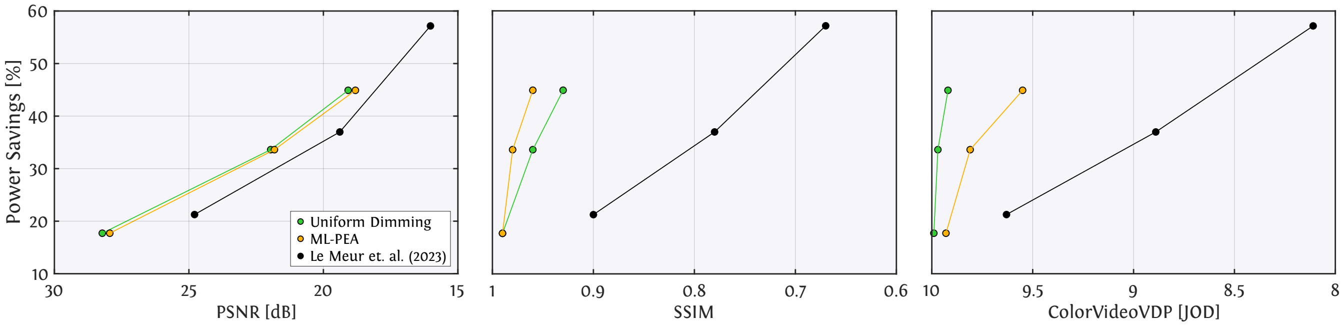

Figure 10: *The results from Table 4 are plotted here, for the three methods compared: uniform dimming (●), ML-PEA (●), and [LMDB23] (●). Quality is on the x-axis and power savings are on the y-axis. Note that here we plot lower quality as x increases, similar to Figure 7.*

10. Quality Metric Correlation Analysis

The typical evaluation strategy for prior machine learning-based display power optimization methods has been to compare the average of metric scores computed across a test dataset. In Section 4.1, we computed quality scores for PSNR, SSIM, and ColorVideoVDP metrics on the DIV2k test dataset, and found that, depending on the quality metric used, the conclusions made about the model's performance are very different. In Figure 11, we computed the root mean square error (RMSE), Spearman (SROCC), Pearson (PLCC), and Kendall (KROCC) correlation coefficients between scores computed by an additional set of metrics (summarized in Table 5) and subjective quality scores from our user study (see Section 5). We recommend PSNR used in prior works should not be used as an evaluation metric as it has a low correlation score, and may not be robust enough for this task.

11. Supplemental Analyses

We conducted a number of additional analyses of the performance of ML-PEA.

11.1. Power Savings Dependency on Image Statistics

In Figure 12, we show that there is a positive correlation between power savings and image statistics. Here, we show the mean and variance of the image. This effect is likely due to the fact that images with many bright regions have greater potential for dimming, and vice versa. In the limit, a completely black image has no room for power savings, for example.

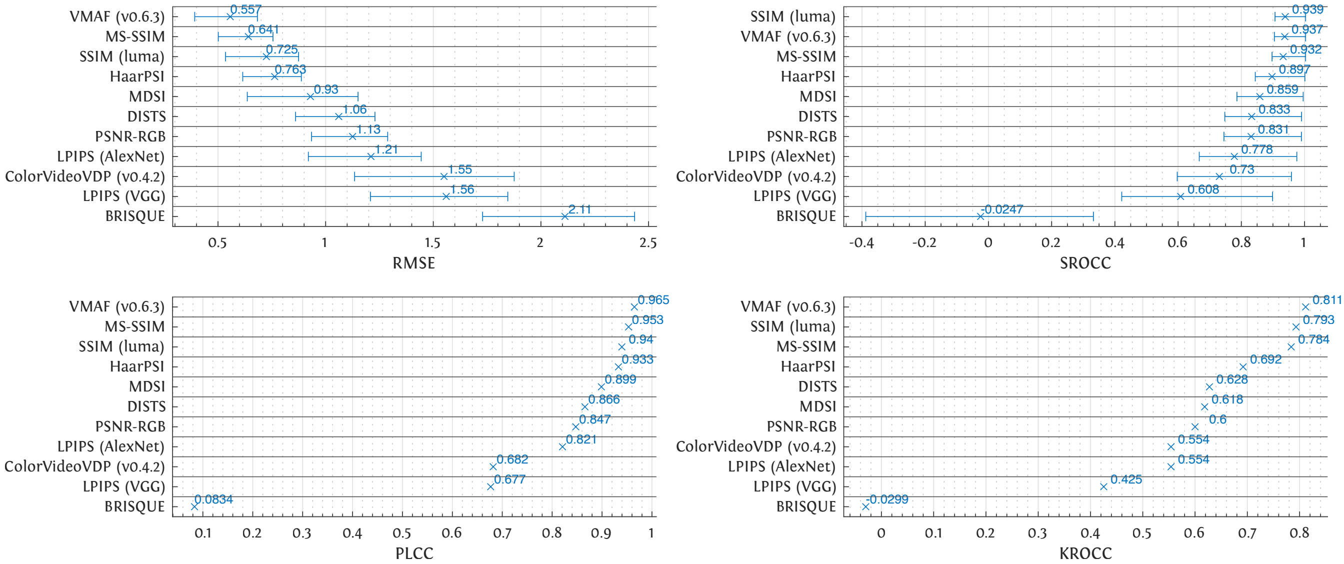

Figure 11: *Correlation results for a number of different quality metrics, ranked by performance.*

Table 5: *Description of quality metrics used in Figure 11.*

Metric	Description
PSNR	Popular metric measuring the ratio between signal and noise.
SSIM [WBSS04]	Quality metric that considers luminance, contrast, and structural differences.
MS-SSIM [WSB03]	Multi-scaled version of SSIM.
LPIPS (VGG) [ZIE*18]	Compares feature representation of images from a pre-trained VGG network.
LPIPS (AlexNet)	Same as LPIPS (VGG) but with an AlexNet backbone.
VMAF [LBN*18]	Perceptual video quality metric that fuses a number of elementary metrics via support vector machines.
HaarPSI [RBKW18]	Perceptual quality measure based on the Haar wavelet decomposition.
MDSI [NSHC16]	Quality metric based on structural and color similarity.
DISTS [DMWS20]	Image quality metric that compares structure and texture similarity using deep features from a pre-trained CNN.
BRISQUE [MMB12]	A no-reference quality metric based on scene statistics.
ColorVideoVDP [MHA*24]	Low-level visual model that considers chromatic and achromatic sensitivity.

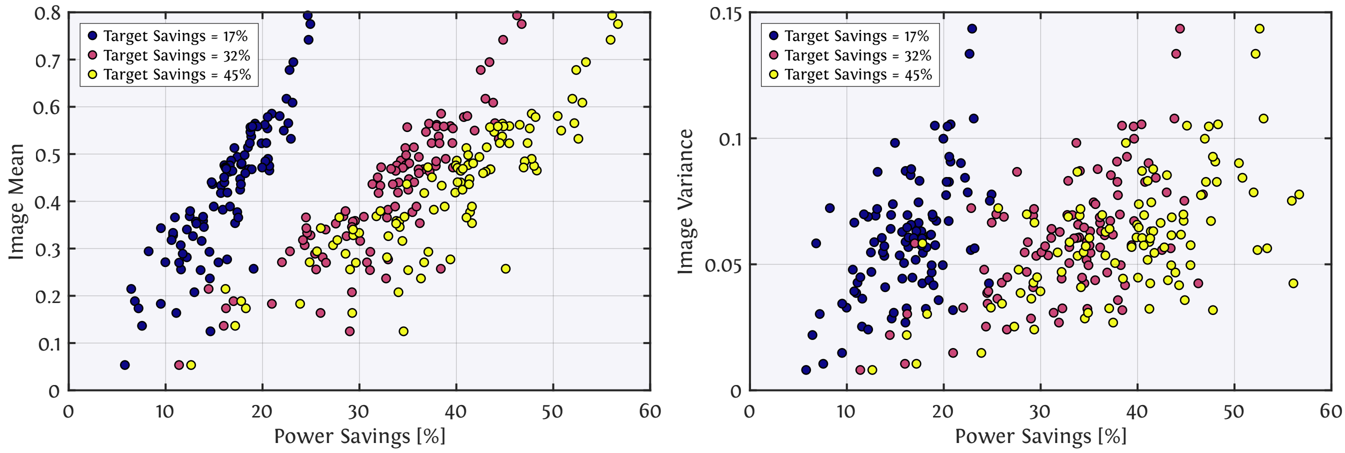

Figure 12: *We show dependence of power savings on image mean (left) and variance (right).*

12. Additional Results

In Figure 13, we show additional results comparing uniform dimming to ML-PEA. The first row shows the input images, and the next rows are power-optimized images at the target power saving rates shown to the left. Zoom in for details.

13. User Study

13.1. Study Instructions

In Figure 14, we show a screen grab of the user study instructions read to the users.

13.2. Just Objectionable Difference

The JOD unit is defined in [POM17]. JODs can be mapped to percentage preference, as shown in Figure 15. They are scaled in a way such that 1 JOD between some condition A and another B equals a percentage selection of A of 75% over B.

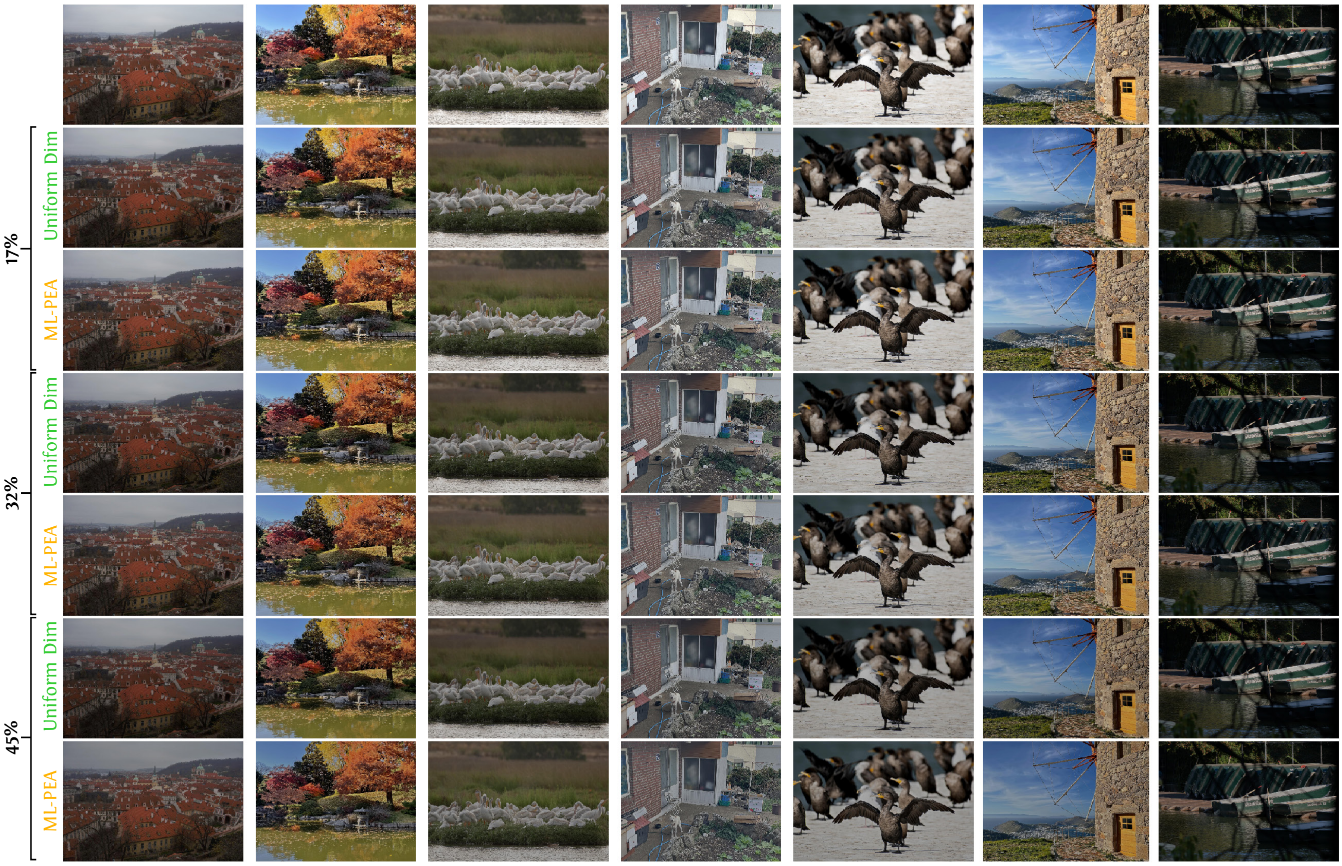

Figure 13: *Additional results are shown here, comparing uniform dimming and ML-PEA.*

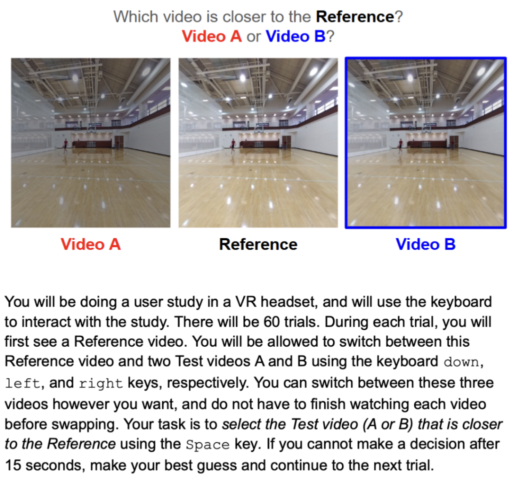

Figure 14: *The study instructions read to users is shown here.*

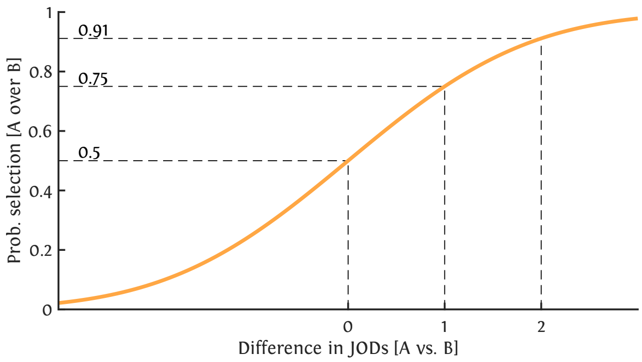

Figure 15: *We map JODs (x-axis) to units of percentage preference (y-axis). Here, we show the probability of selection of a method A over another B for 0, 1, and 2 JODs (50%, 75%, and 91%, respectively).*