Semantic Scale Space: A Framework for Controllable Image Abstraction

Supplementary Material

Supplementary Material Overview

This supplementary document provides additional theory, implementation details, and extended analyses that support the claims made in the main paper. The contents are organized as follows:

- **Section A: Extended Theory.** We provide a deeper theoretical analysis of the Semantic Scale Space (SSS) and AGSS, clarifying the relationship between the classical symmetric Dirichlet energy and our donor-gated formulation, and justifying the latter as both a local surrogate minimizer and a stable non-reversible diffusion operator.
- **Section B: Effect-Matching Protocol Details.** We detail the method-independent Region Homogeneity Index (RHI), the generation of target levels, and the reachability diagnostics used for excluded baselines such as L_0 smoothing and FSPIS.
- **Section C: Ablations and Variants.** We compare donor-gated and symmetric averaging, analyze sensitivity to weaker boundary cues through a Canny-based ablation, and provide additional implementation details and discussion for Semantic Filtering (DT+MuGE).
- **Section D: Sanity Checks (P1, P2, P4).** We provide empirical verification of continuity, monotonicity, and convergence through RHI trajectories and MAD diagnostics.
- **Section E: Granularity Parameter Analysis.** We analyze the behavior of MuGE's granularity parameter α, including edge-density trends and approximate monotonicity over the extended range used in this work.
- **Section F: Runtime Analysis.** We report same-hardware CPU and GPU runtimes, together with an AGSS component breakdown.
- **Section G: User Study Details.** We provide the detailed protocol, statistical analysis, and all stimuli used in the 2AFC user study.

A. Extended Theory of SSS and AGSS

This section provides additional theoretical context for our Semantic Scale Space (SSS) framework and the AGSS operator. Our goal is not to claim a full variational or PDE-equivalent derivation of AGSS, but rather to clarify how the proposed update can be understood as a selective, stable smoothing process related to classical diffusion ideas. We first review a continuum motivation, then discuss a symmetric discrete baseline, and finally justify the donor-gated update from two complementary viewpoints: as the exact minimizer of a local separable quadratic objective, and as a non-reversible row-stochastic averaging operator. We then relate these observations qualitatively to the SSS properties P1, P2, and P4, with empirical diagnostics given in later sections.

A.1. From PDE to Discrete Energies

As motivated in the main paper, a natural starting point for semantic abstraction is the anisotropic diffusion PDE

$$\frac{\partial I}{\partial t} = \nabla \cdot \big(D_g(x)\,\nabla I\big), \tag{S1}$$

where strength t is diffusion time and granularity g controls the conductance field $D_g(x)$. Intuitively, $D_g(x)$ is large within homogeneous semantic regions (allowing diffusion) and small near detected boundaries (suppressing diffusion), and is derived from a boundary likelihood map S_g produced by an external detector.

Formally, (S1) is associated with a semantic Dirichlet energy of the form

$$\mathcal{E}_g(I) \;\propto\; \int D_g(x)\,\|\nabla I(x)\|^2\,dx.$$

In this sense, the PDE provides a useful conceptual reference: it suggests that a desirable abstraction operator should be edge-aware through D_g or S_g, smoothing within regions while suppressing diffusion across important boundaries.

Our method does *not* attempt to solve (S1) exactly. Instead, we use the PDE and its associated energy only as motivation for constructing a more selective discrete operator. In particular, we seek an update rule that remains local and smoothing-like, but is less prone than symmetric discretizations to cross-boundary mixing near salient structures.

A.2. Symmetric Discretization Baseline

A standard discrete analogue of the semantic Dirichlet energy on a pixel grid is the convex quadratic form

$$\mathcal{E}_g^{\mathrm{sym}}(I) = \frac{1}{2}\sum_{x}\sum_{y\in\mathcal{N}_8(x)} w_g^{\mathrm{sym}}(x,y)\,\|I(x)-I(y)\|^2, \tag{S2}$$

where $\mathcal{N}_8(x)$ is the 8-connected neighborhood of pixel x. Including the center pixel x itself in the inner sum would not change the value of $\mathcal{E}_g^{\mathrm{sym}}(I)$, because the term with $y = x$ vanishes as $\|I(x) - I(x)\|^2 = 0$. For convenience, however, we will later use the slightly extended stencil $\mathcal{N}_8^+(x) = \{x\} \cup \mathcal{N}_8(x)$ when defining our donor-gated averaging operator.

A natural *symmetric* edge-aware weight is

$$w_g^{\mathrm{sym}}(x,y) = \eta(x,y)\,\frac{A_g(x)+A_g(y)}{2}, \tag{S3}$$

where $\eta(x,y) \geq 0$ encodes spatial connectivity and $A_g(\cdot) = 1 - S_g(\cdot)$ is a semantic pass weight derived from the boundary likelihood map S_g. This choice treats both endpoints of a connection symmetrically and is consistent with classical edge-weighted Laplacian constructions.

The stationarity condition of (S2), or equivalently a Jacobi relaxation step for the associated linear system, yields updates of the form

$$I^{(n+1)}(x) = \frac{\sum_{y\in\mathcal{N}_8(x)} w_g^{\mathrm{sym}}(x,y)\, I^{(n)}(y)}{\sum_{y\in\mathcal{N}_8(x)} w_g^{\mathrm{sym}}(x,y)},$$

where we use n to denote the discrete iteration index. These are symmetric weighted averages. They are theoretically appealing and closely connected to (S1), but in practice they still allow information exchange in both directions across nearby pixels. As a result, even when A_g is small near detected boundaries, moderate smoothing can visibly soften structural boundaries through residual cross-boundary mixing. This motivates the more selective one-sided design used in AGSS.

A.3. Donor-Gated Averaging: A Local Quadratic Objective and a Non-Reversible Averaging Process

AGSS replaces the symmetric coupling above with a *donor-gated* update that depends only on the semantic pass weight at the value-providing pixel (the donor). We now provide two complementary ways to interpret this operator.

Viewpoint 1: Local separable quadratic objective. Rather than directly minimizing the symmetric energy $\mathcal{E}_g^{\mathrm{sym}}(I)$, we consider, at a given iterate $I^{(n)}$, a simple separable quadratic objective that uses the same spatial weights and donor gates while treating $A_g(y)$ and $I^{(n)}(y)$ as fixed coefficients. For each pixel x, using the stencil $\mathcal{N}_8^+(x) = \{x\} \cup \mathcal{N}_8(x)$, define

$$\begin{aligned}\mathcal{Q}(I \mid I^{(n)}) = \tfrac{1}{2}\sum_x \Big(\sum_{y\in\mathcal{N}_8^+(x)} \eta(x,y)\, A_g(y)\Big)\, \|I(x)\|^2 \\ - \sum_x \sum_{y\in\mathcal{N}_8^+(x)} \eta(x,y)\, A_g(y)\, \langle I(x), I^{(n)}(y)\rangle \\ + C,\end{aligned} \tag{S4}$$

where $\langle\cdot,\cdot\rangle$ denotes the standard inner product over color channels and C collects terms that do not depend on I.

A key feature of $\mathcal{Q}$ is that it is separable over pixels: for each fixed x, the dependence on $I(x)$ is purely quadratic and does not couple different spatial locations. Taking the derivative of $\mathcal{Q}$ with respect to $I(x)$ and setting it to zero gives

$$\begin{aligned}\frac{\partial \mathcal{Q}}{\partial I(x)} = \Big(\sum_{y\in\mathcal{N}_8^+(x)} \eta(x,y)\, A_g(y)\Big)\, I(x) \\ - \sum_{y\in\mathcal{N}_8^+(x)} \eta(x,y)\, A_g(y)\, I^{(n)}(y) = 0,\end{aligned} \tag{S5}$$

so the minimizer satisfies

$$I(x) = \frac{\sum_{y\in\mathcal{N}_8^+(x)} \eta(x,y)\, A_g(y)\, I^{(n)}(y)}{\sum_{y\in\mathcal{N}_8^+(x)} \eta(x,y)\, A_g(y)}. \tag{S6}$$

This is already a donor-gated averaging step: the new value at x is a normalized combination of the previous donor values $I^{(n)}(y)$, weighted only by spatial connectivity and donor-side pass weights.

In practice, AGSS adds a small stabilizer $\xi > 0$. This can be interpreted as augmenting $\mathcal{Q}$ with a Tikhonov-type term $\frac{1}{2}\xi \sum_x \|I(x) - I^{(n)}(x)\|^2$, which biases the update slightly toward the previous iterate and guarantees a strictly positive denominator. The resulting optimality condition is

$$\begin{aligned}\Big(\sum_{y\in\mathcal{N}_8^+(x)} \eta(x,y)\, A_g(y) + \xi\Big)\, I(x) \\ = \sum_{y\in\mathcal{N}_8^+(x)} \eta(x,y)\, A_g(y)\, I^{(n)}(y) + \xi\, I^{(n)}(x),\end{aligned} \tag{S7}$$

and hence

$$I(x) = \frac{\sum_{y\in\mathcal{N}_8^+(x)} \eta(x,y)\, A_g(y)\, I^{(n)}(y) + \xi\, I^{(n)}(x)}{\sum_{y\in\mathcal{N}_8^+(x)} \eta(x,y)\, A_g(y) + \xi}. \tag{S8}$$

Recognizing the numerator and denominator as in the main paper, this update coincides with the donor-gated AGSS step. Thus, AGSS can be interpreted as exactly minimizing a local separable quadratic objective whose coefficients are determined by donor-side semantic pass weights. We emphasize that this is an *interpretive construction*: it clarifies the structure of the update, but it is not claimed to be a global majorizer or exact surrogate of the symmetric Dirichlet energy (S2).

Viewpoint 2: Non-reversible diffusion via a row-stochastic operator. Recall the donor-gated weights $\omega_g(y \to x) = \eta(x,y)\, A_g(y)$ and the normalizer $Z_g(x) = \sum_{y\in\mathcal{N}_8^+(x)} \omega_g(y \to x) + \xi$. For a fixed granularity stage, define a matrix W_g by

$$W_g(x,y) = \begin{cases} \dfrac{\omega_g(y \to x)}{Z_g(x)}, & y \neq x, \\ \dfrac{\omega_g(x \to x) + \xi}{Z_g(x)}, & y = x. \end{cases}$$

Then the update can be written as

$$I^{(n+1)}(x) = \sum_y W_g(x,y)\, I^{(n)}(y).$$

By construction, $W_g(x,y) \geq 0$ and $\sum_y W_g(x,y) = 1$, so W_g is row-stochastic. Consequently, each new pixel value is a convex combination of previous values. This immediately implies several basic stability properties.

First, the update is L^∞**-stable**: for each channel, no value can move outside the previous channel-wise range,

$$\min_x I^{(n)}(x) \;\leq\; I^{(n+1)}(x) \;\leq\; \max_x I^{(n)}(x).$$

Thus the dynamic range cannot be expanded by a single update.

Second, the mapping is **non-expansive in** L^∞: for any two images $I^{(n)}$ and $J^{(n)}$,

$$\|I^{(n+1)} - J^{(n+1)}\|_\infty \;\leq\; \|I^{(n)} - J^{(n)}\|_\infty.$$

This is the standard smoothing behavior of a Markov-type averaging operator.

Third, because the operator is generally *not symmetric*, the associated process is naturally interpreted as a *non-reversible* diffusion or averaging process. The donor-side gating biases propagation away from low-pass donors near boundaries, which is precisely the mechanism used by AGSS to reduce cross-boundary mixing.

For each fixed granularity stage, if the underlying graph is sufficiently connected and the self-loops induced by ξ make the chain aperiodic, repeated application of W_g approaches a stage-wise steady state for each channel. In AGSS, however, the granularity changes across stages, so the full process is better viewed as a sequence of such stage-wise averaging dynamics rather than a single stationary Markov chain.

A.4. Convergence and Monotonic Smoothing (Qualitative Discussion)

The viewpoints above provide qualitative support for the SSS properties targeted in the main paper, but they should be understood as interpretations rather than formal proofs.

P1: Continuity in strength. For fixed granularity g, the donor-gated update is a linear iteration $I^{(n+1)} = W_g I^{(n)}$ with a stable row-stochastic operator W_g. Thus the abstraction evolves through small, repeated averaging steps rather than through abrupt global remappings. Strict continuity is not claimed at the level of discrete iteration, but the process naturally defines a smooth discrete trajectory in strength, consistent with the empirical behavior reported in Section D.

P2: Region-monotonic smoothing. Within non-boundary regions where $A_g(y) \approx 1$, the update behaves like a local convex averaging process. Such averaging typically reduces local variance and suppresses fine-scale texture. A general proof for arbitrary spatially varying donor gates is outside the scope of this work, but the RHI trajectories in Section D show monotonic decrease in practice, supporting P2 empirically.

P4: Groundedness and endpoints. By construction, AGSS starts from the original image, $I^{(0)} = I_0$. For each fixed stage, the update is stable and non-expansive, and under mild connectivity assumptions it approaches a stage-wise steady state. Because AGSS uses a finite sequence of such stages and the per-iteration MAD decreases rapidly within each stage (Section D), the overall process exhibits a well-behaved progression from the input toward a numerically stable simplified image. These observations are consistent with P4.

In summary, AGSS can be interpreted as a selective, stable, non-reversible smoothing process that is inspired by semantic diffusion ideas but deliberately departs from symmetric discretizations in order to better suppress cross-boundary mixing.

B. Effect-Matching Protocol (RHI & Search)

To ensure a fair comparison between methods that operate on different principles, we utilize the Region Homogeneity Index (RHI) to standardize the "amount of smoothing."

B.1. RHI Definition

The RHI quantifies smoothness specifically within non-boundary regions. It is computed as follows:

1. **Preprocessing:** The sRGB input image is converted to linear luminance.
2. **Masking:** A non-boundary mask is created by dilating the ground-truth boundary map (radius 6 pixels) and inverting it. This excludes pixels on or near semantic boundaries from the measurement.
3. **Variance Map:** A local variance map is computed over the linear luminance image using a 7×7 window.
4. **Aggregation:** The RHI is defined as the mean of these local variance values within the non-boundary mask.

This metric provides a method-independent measure of texture suppression that is robust to boundary sharpness.

B.2. Target Generation

To define consistent abstraction levels, we generate Reference RHIs by applying Gaussian blur to the input linear luminance image and measuring the resulting RHI using the procedure above.

Table S1. Reachability diagnostics for excluded baselines under the RHI-matching protocol on SBD. Reach% denotes the fraction of images for which the target level could be reached. BPF-ODS and Drift are computed only on the reached subset.

Method	Level	Reach%	BPF-ODS ↑	Drift ↓
L_0	Weak	98.0	0.442 ± 0.181	99.17 ± 37.15
	Medium	91.5	0.398 ± 0.182	105.83 ± 36.62
	Strong	83.6	0.364 ± 0.194	111.25 ± 39.95
FSPIS	Weak	74.6	0.518 ± 0.176	91.35 ± 37.51
	Medium	14.6	0.601 ± 0.193	85.37 ± 38.71
	Strong	3.4	0.666 ± 0.236	81.74 ± 42.47

- **SBD (Fixed Scale):** For the SBD dataset (E1), we use fixed Gaussian standard deviations $\sigma \in \{1.0, 2.0, 3.0\}$ to define the Weak, Medium, and Strong targets.
- **DIV2K (Relative Scale):** For the DIV2K dataset (E2), image resolutions vary significantly. We therefore use image-size–normalized standard deviations $\sigma_{\text{img}} = \sigma_{\text{rel}}\sqrt{H^2 + W^2}$, with $\sigma_{\text{rel}} \in \{0.006, 0.018\}$ for Weak and Strong targets, respectively. Since ground-truth boundaries are unavailable for DIV2K, the RHI is measured over all pixels.

B.3. Reachability Issues and Excluded Baselines

Under our RHI-matching protocol, some methods fail to reach the target abstraction levels reliably across the dataset. Since our main comparison is effect-matched, such methods are excluded from the main paper and instead analyzed here through reachability diagnostics. The quantitative results are summarized in Table S1.

L_0 smoothing. As also illustrated by the trajectory examples in Section D, L_0 smoothing exhibits unstable behavior under our protocol: increasing the sparsity weight may cause the process to saturate or collapse structure before the desired target is reached. Its reachability remains relatively high at the weak level but decreases as the target becomes stronger.

Deep smoothing with limited controllability. For FSPIS [3], the main issue is limited adjustable range. While some images can be matched at weak levels, reachability drops sharply for stronger targets. This means that the reported metrics at medium/strong levels are computed only on a small, easier subset of images and therefore are not directly comparable to full-dataset results.

The seemingly strong FSPIS scores at medium/strong levels are therefore misleading: they reflect only the small subset of images that are easiest to smooth to the target level. To verify this subset-bias effect, we additionally evaluate AGSS on exactly the same FSPIS-reached subset. The corresponding results are reported in Table S2.

Table S2. AGSS evaluated on the subset of images for which FSPIS reaches the target RHI. AGSS remains superior even on this subset, confirming that the high FSPIS scores in Table S1 are largely driven by subset bias.

Level	BPF-ODS ↑	Drift ↓
Weak	0.534 ± 0.172	88.22 ± 36.31
Medium	0.630 ± 0.178	80.87 ± 33.73
Strong	0.713 ± 0.197	70.55 ± 35.37

These diagnostics clarify why we exclude L_0 smoothing and FSPIS from the main effect-matched comparison: the former does not provide a stable monotonic path to stronger targets, and the latter reaches only a narrow and highly biased subset at medium/strong levels.

B.4. Additional Visual Comparisons for E1

To complement the main paper, we show additional effect-matched visual comparisons on three SBD test images in Figures S1–S3.

C. Ablations and Variants

C.1. Symmetric vs. Donor-Gated Averaging

To isolate the contribution of the donor-gated mechanism derived in Section A.3, we compared AGSS against a variant using symmetric weights:

$$w^{\text{sym}}(x, y) = \eta(x, y)\,\frac{A_g(x) + A_g(y)}{2}.$$

As summarized in Table S3, donor-gating consistently yields higher BPF-ODS at all matched effect levels. Qualitatively, the symmetric variant tends to blur semantic edges, which lowers the RHI before texture is fully removed; as a result, at matched RHI levels it often retains more residual texture than donor-gating. In contrast, donor-gating better preserves sharper boundaries while producing cleaner region flattening.

C.2. Additional Discussion on Semantic Filtering (DT+MuGE)

To complement the main-paper comparison, we provide additional implementation details and discussion for Semantic Filtering [7] using Domain Transform (DT) [2] with MuGE-based boundary cues.

We use MuGE with a fixed granularity setting $\alpha = 1$ and search the DT range parameter σ_r on a validation subset of SBD. While a smaller value ($\sigma_r = 0.1$) yields slightly better boundary scores at weak and medium levels, it fails to reach the strong target on many images because the smoothing range is too limited. We therefore fix $\sigma_r = 0.3$ for all

Figure S1. Effect-matched comparison for E1 on SBD image `2008_000195`. Rows: weak/medium/strong effect levels (matched via RHI). Columns: Gaussian reference at the target RHI, WLS, PM, GF-it, and **AGSS**.

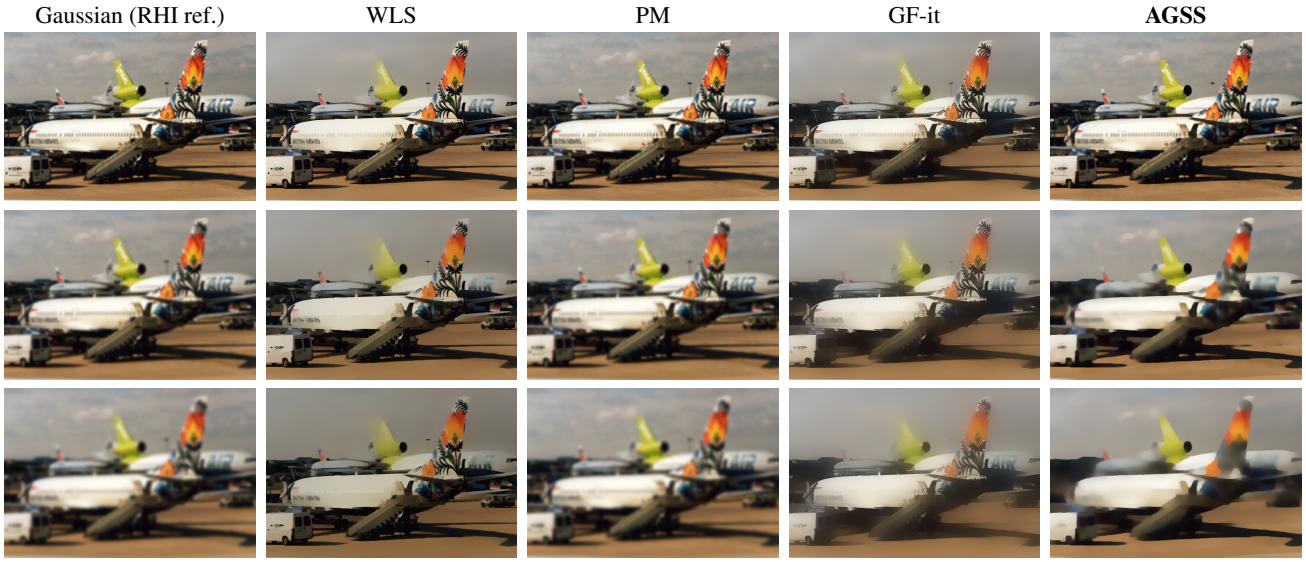


Figure S2. Effect-matched comparison for E1 on SBD image `2008_000251`. Same layout as Fig. S1.

experiments, and tune the DT spatial parameter σ_s per image to match the target RHI.

Table S4 shows that Semantic Filtering improves slightly as the target becomes stronger, reflecting increased texture suppression. However, it remains clearly behind AGSS at all three levels. This indicates that MuGE-based semantic boundary cues alone are insufficient: the way these cues are incorporated into the smoothing operator is also critical.

In particular, DT+MuGE faces an inherent trade-off. To reach strong abstraction levels, σ_r must be set high enough to allow substantial smoothing, but this also increases tolerance to intensity differences and weakens boundary protection near less prominent structures. As a result, boundary-adjacent texture tends to persist or contours become less coherent. In contrast, AGSS combines donor-gating with a staged fine-to-coarse traversal, enabling stronger abstraction while maintaining sharper boundaries and lower drift.

C.3. Sensitivity to Boundary Quality: Canny vs. MuGE

Since AGSS relies on an external boundary detector, a natural question is how its performance changes when semantic

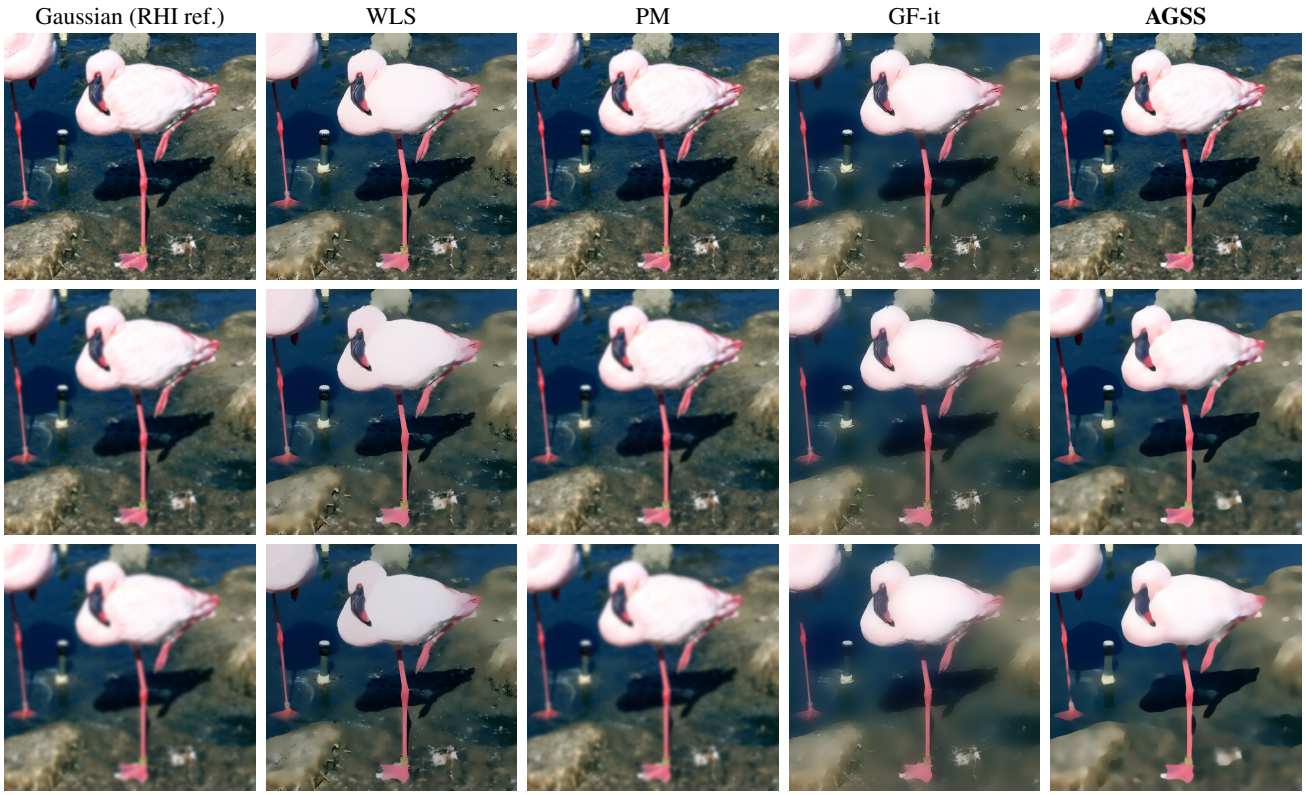


Figure S3. Effect-matched comparison for E1 on SBD image `2008_000515`. Same layout as Fig. S1.

Table S3. Ablation on donor-gated vs. symmetric averaging, evaluated at the three matched RHI levels on SBD. Entries are BPF-ODS (mean ± standard deviation) over the test set.

Averaging method	Weak	Medium	Strong
Symmetric averaging	0.468 ± 0.178	0.508 ± 0.164	0.515 ± 0.163
Donor-gated (AGSS)	$\mathbf{0.535 \pm 0.169}$	$\mathbf{0.575 \pm 0.159}$	$\mathbf{0.594 \pm 0.154}$

Table S4. Effect-matched evaluation of Semantic Filtering (DT+MuGE) on SBD. BPF-ODS is higher-is-better and Drift is lower-is-better.

Level	BPF-ODS ↑	Drift ↓
Weak	0.497 ± 0.176	92.45 ± 36.80
Medium	0.535 ± 0.168	89.04 ± 35.71
Strong	0.555 ± 0.168	88.97 ± 35.88

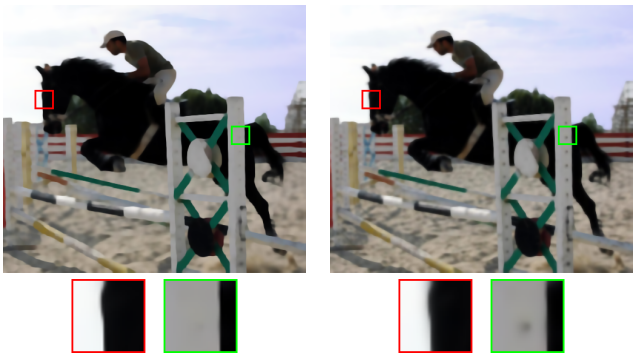
boundary quality degrades. To assess this sensitivity, we replace MuGE with the Canny detector and run the same effect-matched protocol as in E1.

To make this comparison as meaningful as possible, we selected a small set of Canny parameterizations whose outputs visually resemble several MuGE granularity levels on a small set of test images. This matching is only approximate and is intended to provide a weaker, non-semantic boundary source for AGSS rather than an exact replacement for MuGE.

As shown in Table S5, replacing MuGE with Canny causes a clear degradation in both boundary preservation and geometric fidelity. This confirms that semantic boundary quality is important for achieving high selectivity. At the same time, the gap between DT+MuGE and AGSS in the main paper shows that boundary cues alone do not determine performance; the smoothing dynamics also matter.

D. Sanity Checks for SSS Properties

We empirically validate the core SSS properties P1, P2, and P4. We use two SBD images, one "easy" and one "hard" (Figure S5), to analyze the behavior of AGSS and excluded baselines.

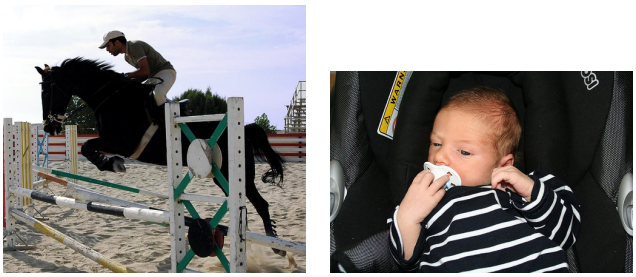


(a) Donor-gated (AGSS, medium level) (b) Symmetric averaging (medium level)

Figure S4. Qualitative comparison of donor-gated and symmetric averaging at the medium abstraction level on a representative SBD image. The red boxed region highlights an edge area and the green boxed region highlights a textured area; both are enlarged below each result. Donor-gating better preserves edge sharpness in the red region and suppresses residual texture more effectively in the green region. The symmetric variant reaches the matched RHI partly by blurring edges, which leaves more residual texture at the same effect level.

Table S5. AGSS with Canny-based boundary cues, evaluated at matched RHI levels on SBD.

Level	BPF-ODS ↑	Drift ↓
Weak	0.473 ± 0.172	92.33 ± 36.71
Medium	0.489 ± 0.162	91.17 ± 34.40
Strong	0.492 ± 0.155	91.36 ± 33.71



(a) 2008_000008 (b) 2008_000213

Figure S5. SBD images used in the RHI-trajectory analysis.

RHI Trajectories (P1 & P2). Figure S6 plots RHI vs. iterations. AGSS shows a strictly monotonic decrease (P2) and smooth evolution (P1). In contrast, FSPIS often cannot reach the target RHI (Figure S7), and L_0 smoothing (Figure S8) exhibits non-monotonic behavior where RHI can increase despite stronger regularization, confirming they do not form a stable scale space across all images.

Convergence (P4). P4 requires that the process starts from the original image and converges toward a simplified state. By construction, AGSS initializes at $I^{(0)} = I_0$. To verify convergence, we analyze two diagnostics:

1. **RHI Saturation:** As iterations proceed, the RHI curves in Figure S6 steadily decrease and eventually saturate at a small positive value, indicating that additional texture suppression becomes negligible. This contrasts with FSPIS and L_0, which may saturate prematurely at higher RHI levels (Figures S7 and S8).
2. **Numerical Stability (MAD):** We also track the per-iteration Mean Absolute Difference (MAD). As shown in Figure S9, ΔMAD decays exponentially within each granularity stage. The transient spikes correspond exactly to granularity switches (α updates), followed by immediate re-stabilization.

These observations jointly confirm that AGSS is grounded and reliably converges to a stable, simplified endpoint.

E. Behavior of MuGE Parameter α

Standard usage of MuGE [8] typically limits $\alpha \in [0, 1]$. However, for abstraction tasks, "over-segmentation" is often desirable in the early stages to protect fine textures that might technically be boundaries. We explored an extended range $\alpha \in [0, 5]$ to serve as a continuous control for semantic density.

E.1. Edge Density Analysis

To quantify the global effect of α, we define edge density as the mean value of the MuGE edge likelihood map $S_\alpha(x)$ across all pixels x. As shown in Figure S10, edge density increases monotonically with α. Figure S12 visually confirms that $\alpha > 1$ captures extremely fine stroke-like details. This justifies our schedule $G = [3, 2, 1, 0]$, which effectively utilizes this extended "super-fine" granularity to initialize the abstraction process conservatively before coarsening.

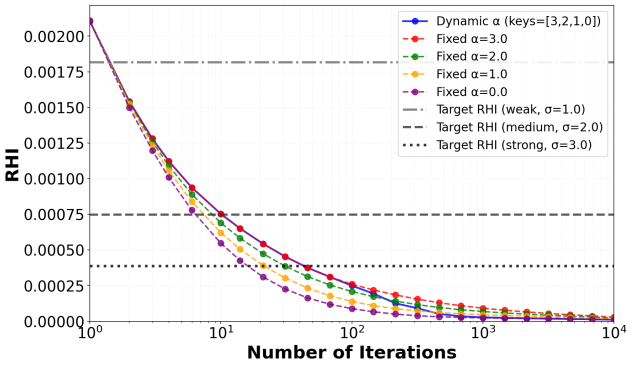
E.2. Monotonicity Analysis

Ideally, increasing α should yield non-decreasing edge likelihoods at each pixel. While strict pixel-wise monotonicity is not guaranteed by the network, we evaluate the stability of α by measuring local violations. We define a violation at pixel x between adjacent granularity steps $\alpha_i < \alpha_{i+1}$ if the edge likelihood decreases by more than a tolerance threshold $\epsilon_{\text{mono}} = 0.05$ (i.e., $S_{\alpha_{i+1}}(x) < S_{\alpha_i}(x) - \epsilon_{\text{mono}}$). Figure S11 confirms that such violations are negligible (low rate and magnitude), ensuring predictable behavior even in this extended range.

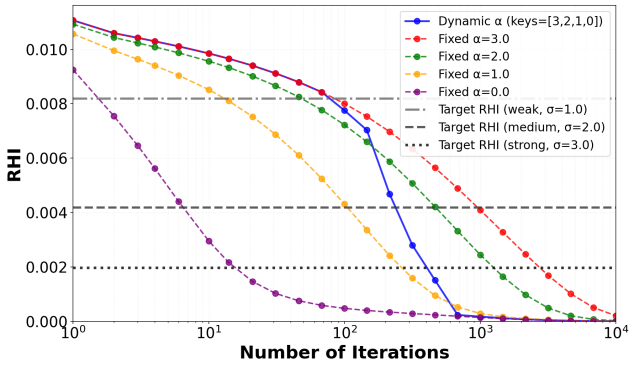
F. Runtime Analysis

F.1. Measurement Setup

To evaluate computational cost under consistent conditions, we measured runtimes on the 18 DIV2K images used in the user study (15.60 MPix in total), resized to 768 px on the

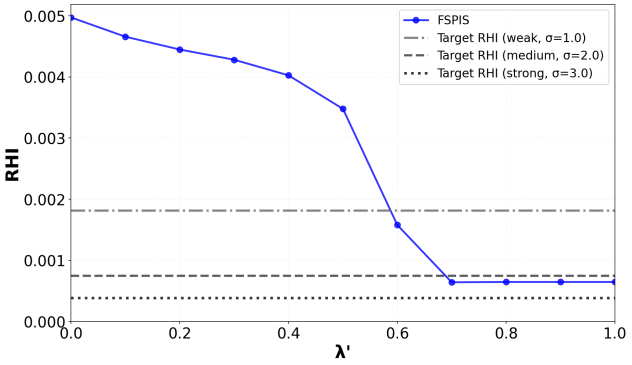


(a) `2008_000008` (Easy).

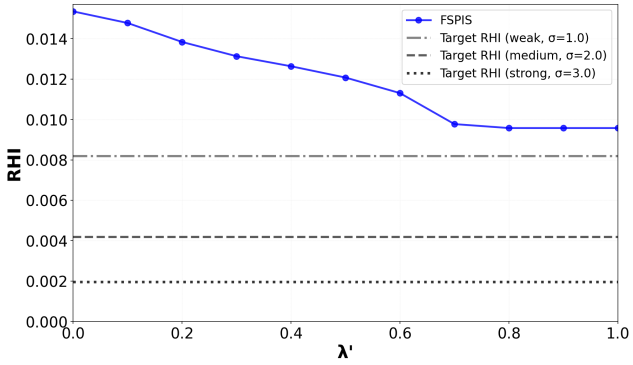


(b) `2008_000213` (Hard).

Figure S6. RHI trajectories of AGSS vs. iterations. RHI decreases smoothly and monotonically, crossing all target levels.



(a) `2008_000008`: Strong target unreachable.



(b) `2008_000213`: Weak target unreachable.

Figure S7. RHI trajectories of FSPIS. The limited range makes matching targets impossible for many images.

short edge. All methods were run on the same workstation in two settings: CPU-only and GPU-only. The CPU was an Intel Core i9-13900KF processor and the GPU was an NVIDIA GeForce RTX 3090. We report runtime per megapixel (s/MPix) for the weak and strong abstraction levels.

F.2. Results on CPU

Table S6 reports CPU-only runtimes. On CPU, AGSS is slower than PM, WLS, and weak-level GF-it, largely because MuGE inference dominates the total cost. The cost further increases at the strong level due to additional smoothing iterations.

Table S6. CPU runtime per megapixel (s/MPix).

Method	Weak [s/MPix]	Strong [s/MPix]
WLS	10.66	10.78
GF-it	7.64	15.14
PM	0.85	3.27
AGSS	15.55	24.69

F.3. Results on GPU

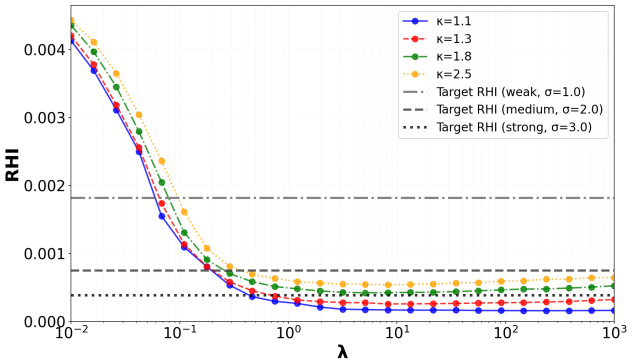
Table S7 reports GPU-only runtimes. On GPU, AGSS runs at 0.59 s/MPix (weak) and 1.06 s/MPix (strong), making it competitive with the strongest baselines while retaining the controllable multi-stage abstraction behavior of AGSS.

Table S7. GPU runtime per megapixel (s/MPix).

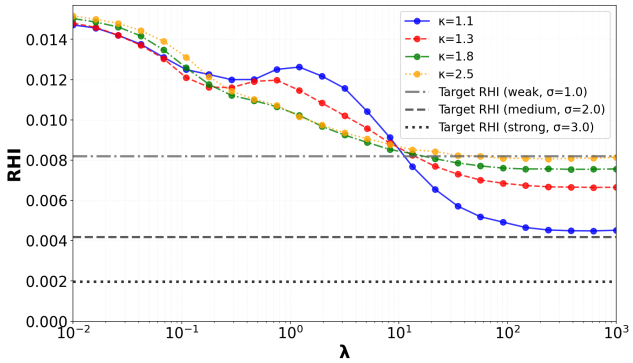
Method	Weak [s/MPix]	Strong [s/MPix]
WLS	0.21	0.21
GF-it	0.54	1.38
PM	0.03	0.11
AGSS	0.59	1.06

F.4. AGSS Breakdown

Table S8 decomposes the AGSS runtime into boundary detection (MuGE) and smoothing. On GPU, MuGE dominates the weak-level runtime, while smoothing becomes comparable at the strong level. On CPU, MuGE dominates both levels, explaining the larger total runtime.

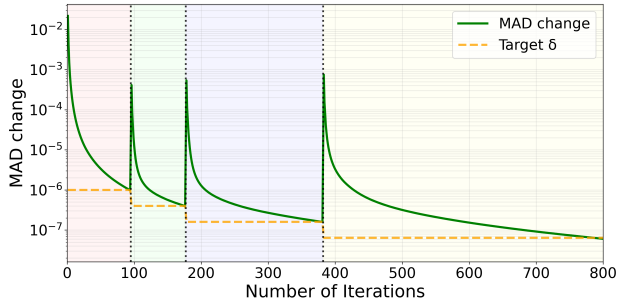


(a) 2008_000008: Non-monotonic behavior.

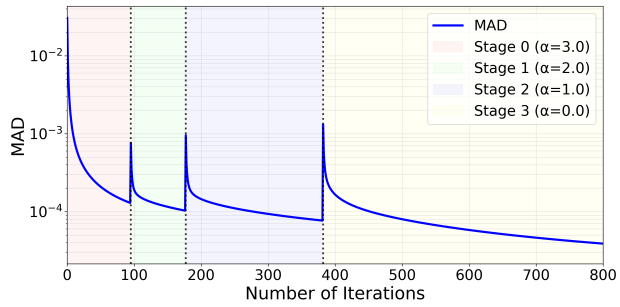


(b) 2008_000213: Saturates above targets.

Figure S8. RHI trajectories of L_0 smoothing. The process is not strictly monotonic and often saturates before reaching the desired abstraction level.



(a) $\Delta\mathrm{MAD}(t)$.



(b) $\mathrm{MAD}(t)$.

Figure S9. Per-iteration MAD diagnostics for AGSS. Spikes occur only at stage transitions (dotted lines), followed by rapid convergence.

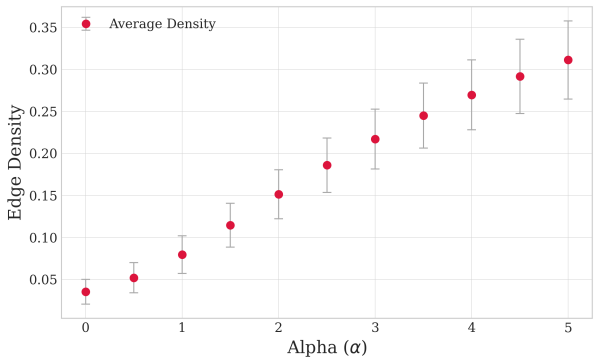


Figure S10. MuGE parameter α vs. edge density. Density increases monotonically with α.

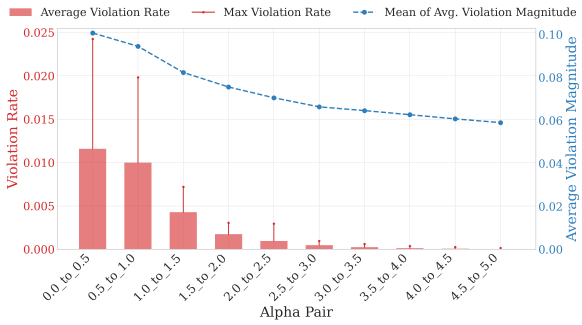


Figure S11. Monotonicity violation analysis. Violations are rare and small in magnitude.

F.5. Implications for Interactivity

The progressive nature of AGSS remains advantageous for interactive control. Unlike global solvers such as WLS, which require a fresh solve when the strength changes, AGSS naturally produces an ordered abstraction trajectory. This allows users to stop at the desired point along the trajectory and obtain immediate visual feedback, especially in the GPU setting where the per-megapixel cost is close to interactive use.

G. User Study Details (2AFC)

We conducted a Two-Alternative Forced-Choice (2AFC) user study to evaluate whether the selectivity gains of AGSS translate into perceptually superior results in downstream NPR applications.

(a) Original Image (b) $\alpha = 0$ (c) $\alpha = 1$ (d) $\alpha = 3$ (e) $\alpha = 5$

Figure S12. Visual effect of increasing MuGE granularity α. Values > 1 capture fine textures.

Table S8. AGSS runtime breakdown (s/MPix).

Setting	Component	Weak	Strong
CPU	MuGE (edges)	14.51	14.87
	Smoothing	1.03	9.82
	Total	**15.55**	**24.69**
GPU	MuGE (edges)	0.53	0.53
	Smoothing	0.06	0.53
	Total	**0.59**	**1.06**

G.1. Stimuli and Conditions

The study used a fixed set of 18 diverse images from the DIV2K dataset [1]. To ensure robust evaluation across different styles, the images were partitioned into three disjoint sets of six images (Set A, B, C), each permanently assigned to one of three downstream NPR pipelines:

- **Set A:** XDoG line art [6].
- **Set B:** White-box Cartoonization (WBC) [5].
- **Set C:** Artistic Sketching [4].

All images were isotropically resized to a short edge of 768 pixels to standardize the spatial scale.

For each image, we generated "Weak" and "Strong" pre-abstracted bases using AGSS and three baselines (WLS, PM, GF-it). Crucially, the abstraction levels were strictly matched using the RHI protocol (Sec. B). The assigned NPR pipeline was then applied to the original image, the Weak base, and the Strong base. These three outputs formed a *stylization sequence* (Original $\rightarrow$ Weak $\rightarrow$ Strong) representing a progression of simplification.

G.2. Experimental Design

The experiment employed a within-subjects design consisting of three blocks (one per NPR pipeline).

- **Counterbalancing:** The presentation order of the three blocks was counterbalanced across participants (using all 6 permutations) to mitigate order effects.
- **Baseline Balancing:** Within each block of 6 trials, AGSS was compared against each baseline (GF-it, WLS, PM) exactly twice. The assignment of baselines to specific images was randomized.
- **Randomization:** On each trial, the left/right position of the AGSS sequence was randomized with equal probability ($p = 0.5$) to prevent positional bias.

G.3. Procedure and Interface

Participants accessed the study via a custom web interface.

1. **Training:** Users completed two practice trials to familiarize themselves with the layout and task.
2. **Main Task:** For each of the 18 trials, participants viewed two columns of images. Each column displayed the stylization sequence (Original, Weak, Strong) for a single method.
3. **Instruction:** The prompt was: *"Decide which image set looks more naturally simplified, and select that side."*
4. **Interaction:** Participants clicked a column to select it and pressed "Next" to confirm. No time limit was imposed.

G.4. Statistical Analysis

A total of 24 participants completed the study, yielding $24 \times 18 = 432$ valid responses. We calculated the preference rate for AGSS (the proportion of trials where AGSS was chosen). Significance was determined using two-sided binomial tests against a random chance level of 50%, and error bars denote 95% Wilson score confidence intervals.

G.5. Visual Examples of Stimuli

Figures S13–S18 show all stimuli used in the 2AFC study, grouped by NPR pipeline and image set.

References

[1] Eirikur Agustsson and Radu Timofte. NTIRE 2017 Challenge on Single Image Super-Resolution: Dataset and Study. In *IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, pages 1122–1131, 2017. 10

[2] Eduardo S.L. Gastal and Manuel M. Oliveira. Domain transform for edge-aware image and video processing. *ACM Transactions on Graphics*, 30(4):69:1–69:12, 2011. 4

[3] Mingjia Li, Yuanbin Fu, Xinhui Li, and Xiaojie Guo. Deep Flexible Structure Preserving Image Smoothing. In *Proceedings of the 30th ACM International Conference on Multimedia*, pages 1875–1883, New York, NY, USA, 2022. Association for Computing Machinery. 4

[4] Zhengyan Tong, Xuanhong Chen, Bingbing Ni, and Xiaohang Wang. Sketch Generation with Drawing Process Guided by

Vector Flow and Grayscale. *Proceedings of the AAAI Conference on Artificial Intelligence*, 35(1):609–616, 2021. 10

[5] Xinrui Wang and Jinze Yu. Learning to Cartoonize Using White-Box Cartoon Representations. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 8087–8096, 2020. 10

[6] Holger Winnemöller, Jan Eric Kyprianidis, and Sven C. Olsen. XDoG: An eXtended difference-of-Gaussians compendium including advanced image stylization. *Computers & Graphics*, 36(6):740–753, 2012. 10

[7] Qingxiong Yang. Semantic Filtering. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 4517–4526, 2016. 4

[8] Caixia Zhou, Yaping Huang, Mengyang Pu, Qingji Guan, Ruoxi Deng, and Haibin Ling. MuGE: Multiple Granularity Edge Detection. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 25952–25962, 2024. 7

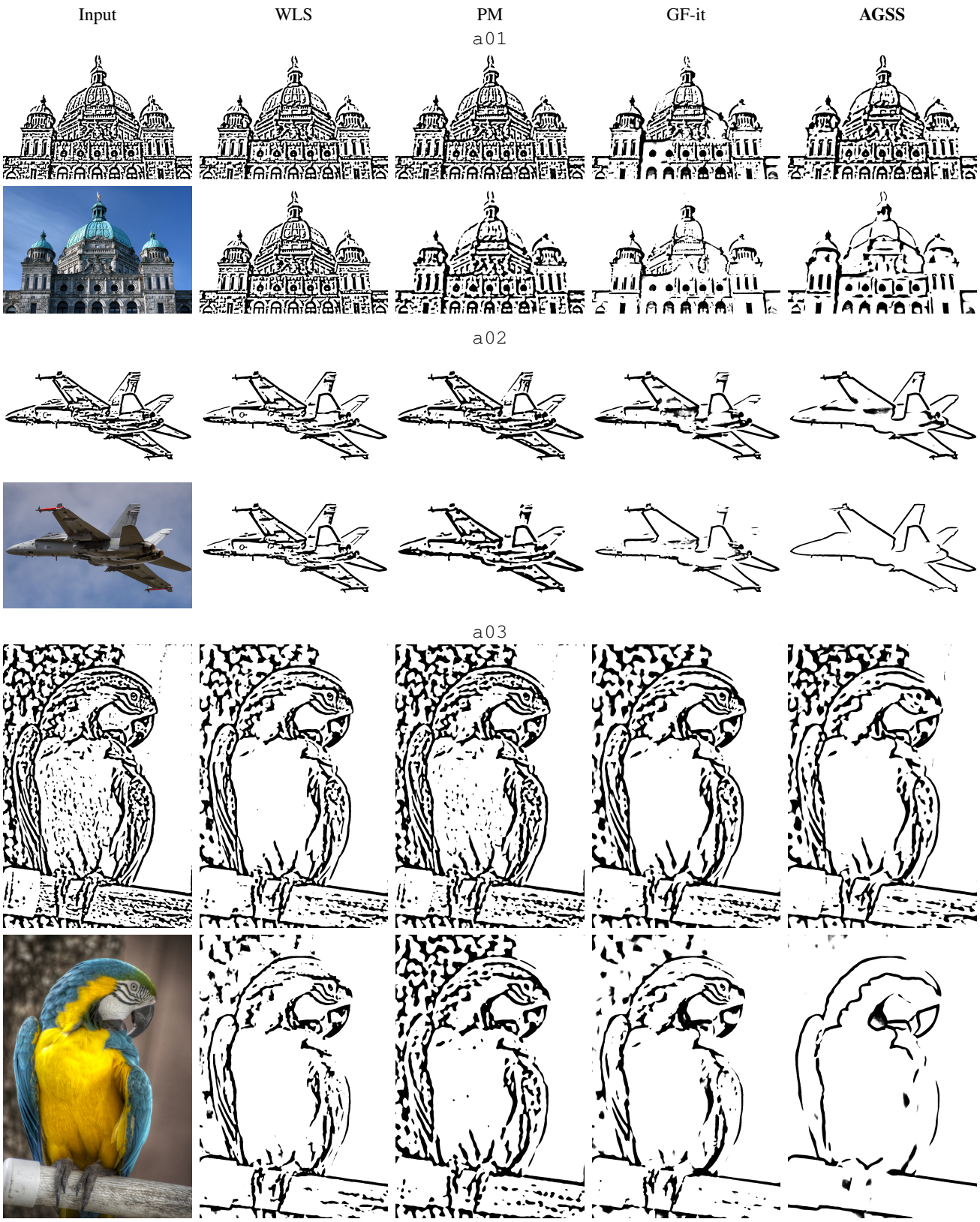


Figure S13. All user-study stimuli for the XDoG pipeline (Set A), examples a01–a03. For each example, the top row shows weak pre-abstraction and the bottom row shows strong pre-abstraction. In the leftmost column, the top image is XDoG applied directly to the original image and the bottom image is the original photograph.

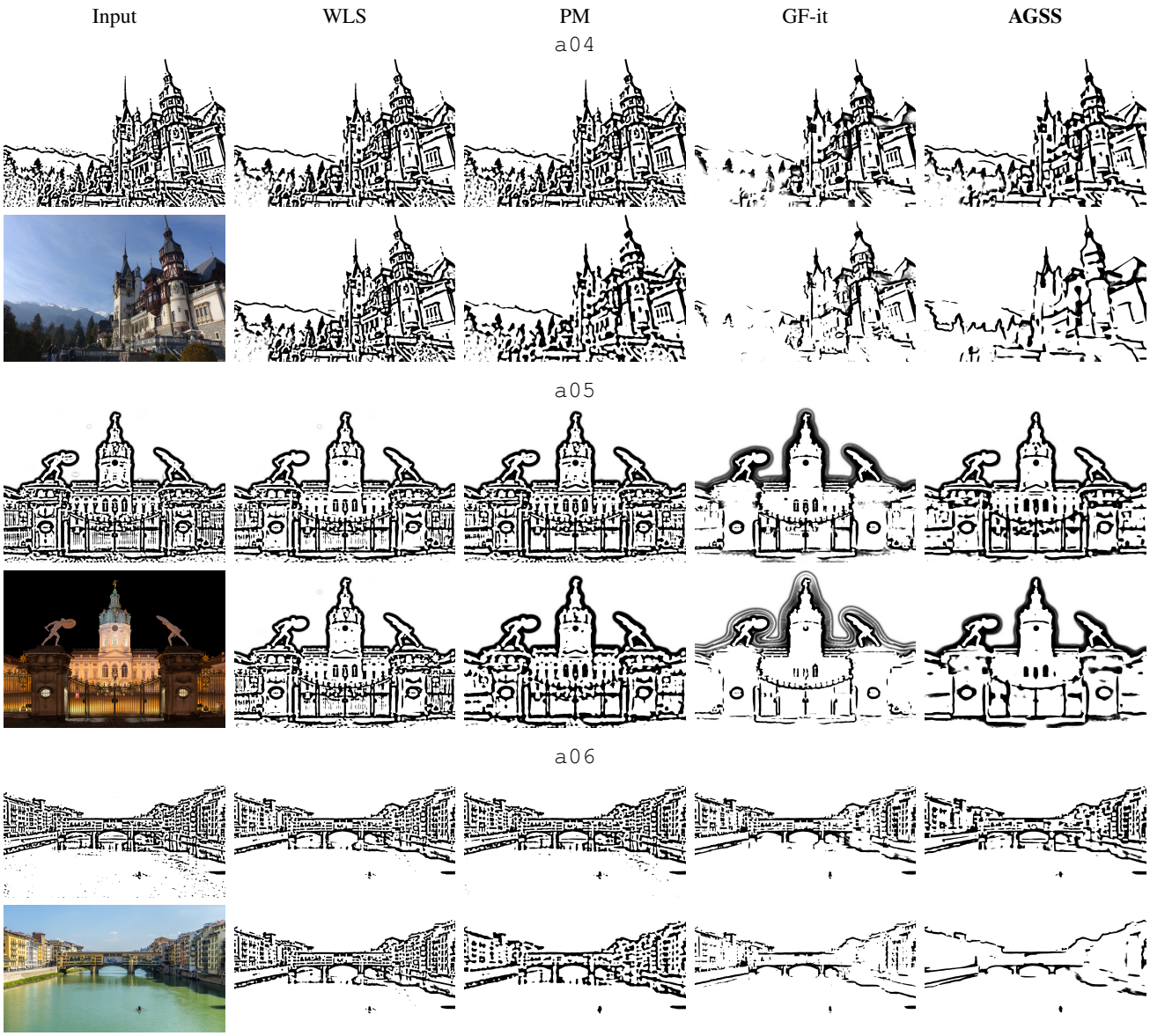


Figure S14. All user-study stimuli for the XDoG pipeline (Set A), examples a04–a06. For each example, the top row shows weak pre-abstraction and the bottom row shows strong pre-abstraction. In the leftmost column, the top image is XDoG applied directly to the original image and the bottom image is the original photograph.

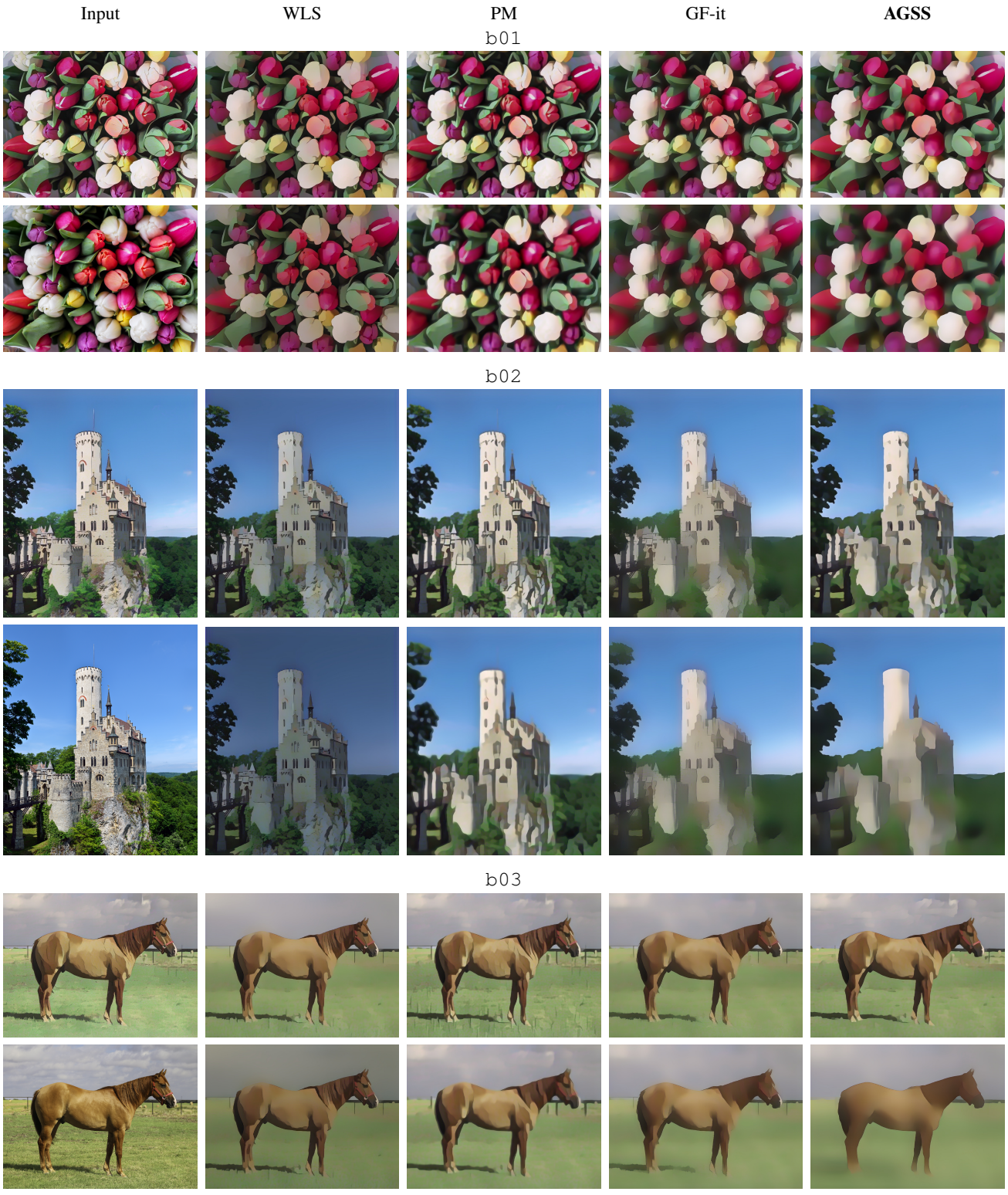


Figure S15. All user-study stimuli for the WBC pipeline (Set B), examples b01–b03. For each example, the top row shows weak pre-abstraction and the bottom row shows strong pre-abstraction. In the leftmost column, the top image is WBC applied directly to the original image and the bottom image is the original photograph.

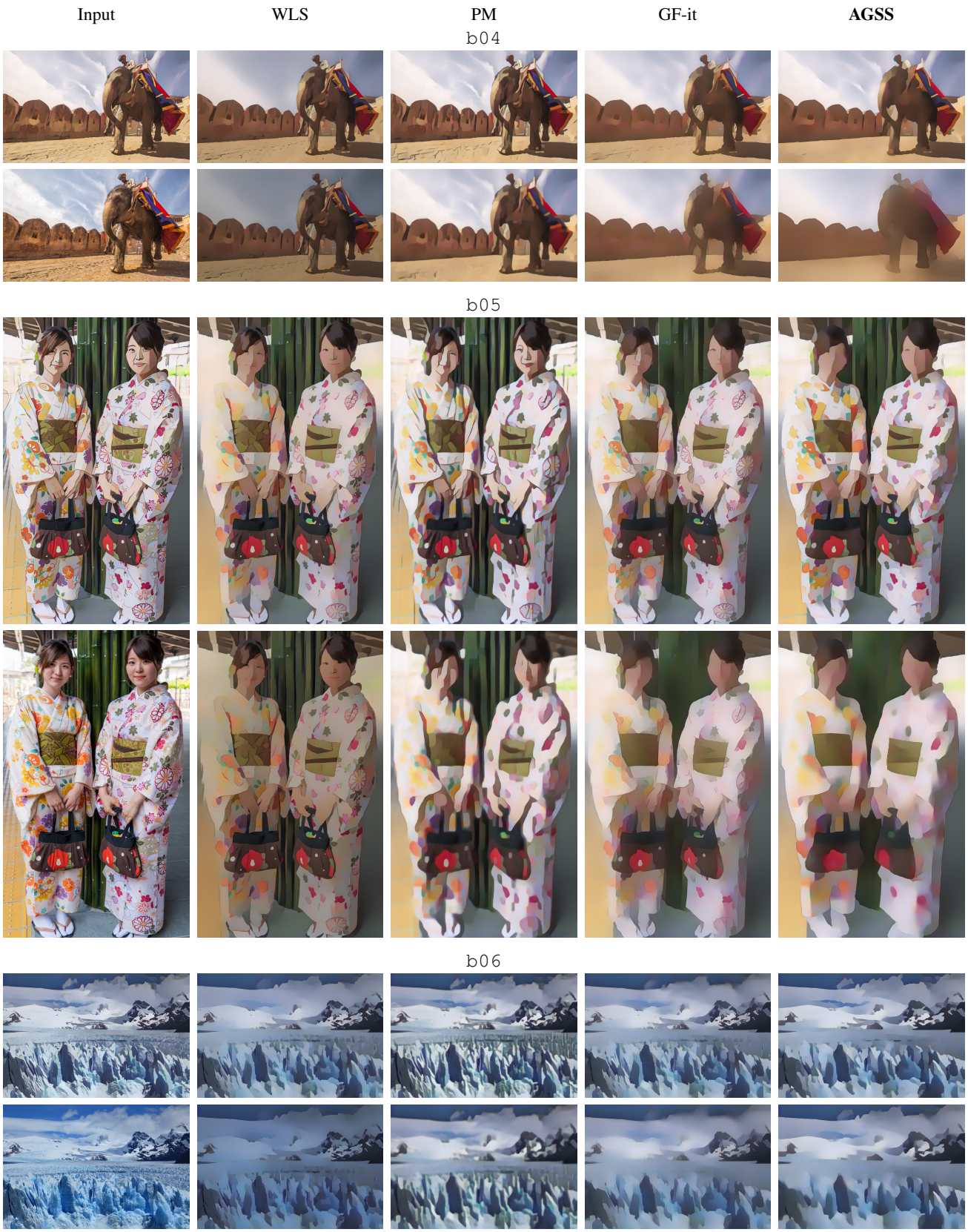


Figure S16. All user-study stimuli for the WBC pipeline (Set B), examples b04–b06. For each example, the top row shows weak pre-abstraction and the bottom row shows strong pre-abstraction. In the leftmost column, the top image is WBC applied directly to the original image and the bottom image is the original photograph.

Figure S17. All user-study stimuli for the Sketch pipeline (Set C), examples c01–c03. For each example, the top row shows weak pre-abstraction and the bottom row shows strong pre-abstraction. In the leftmost column, the top image is Sketch applied directly to the original image and the bottom image is the original photograph.

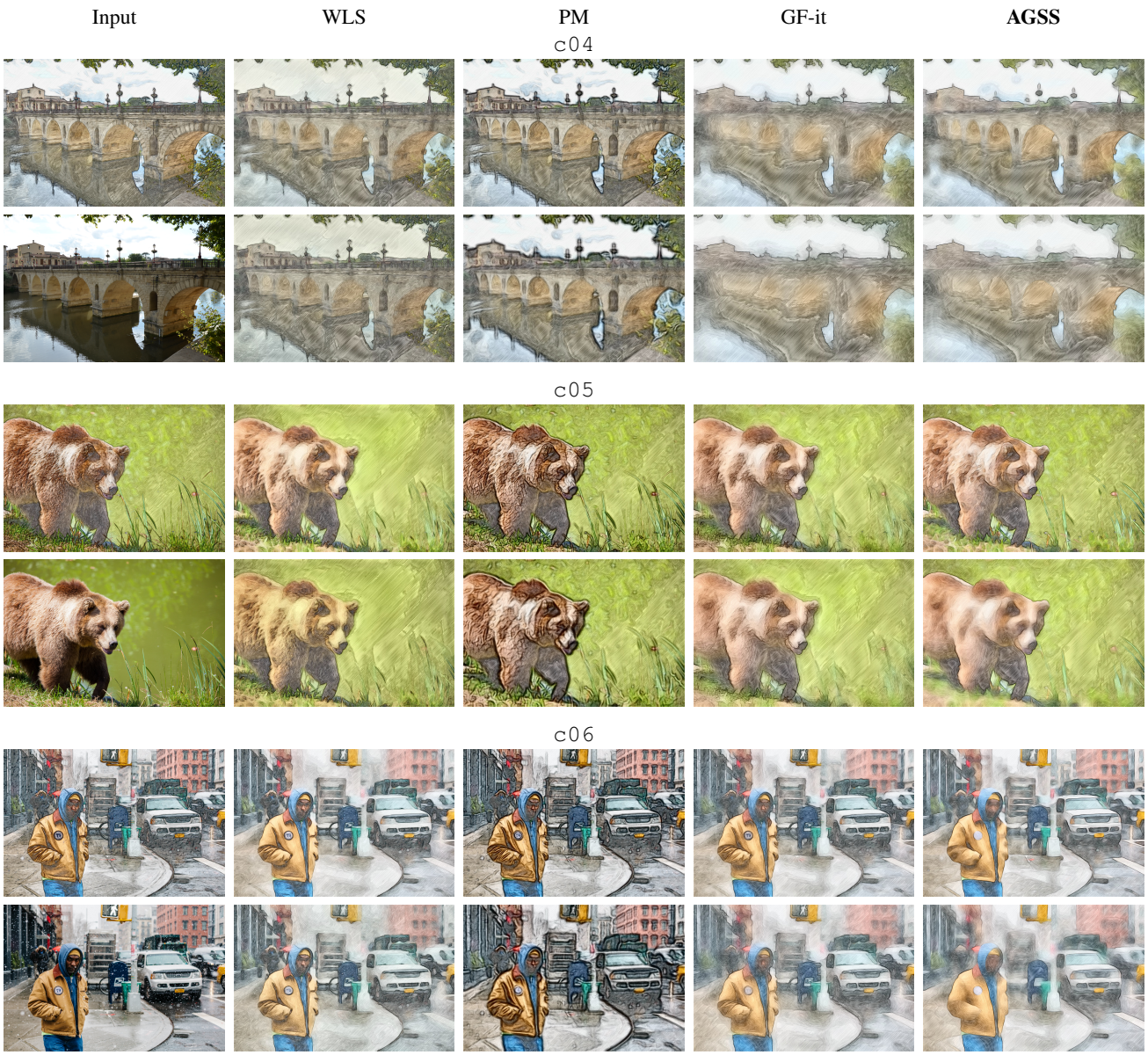


Figure S18. All user-study stimuli for the Sketch pipeline (Set C), examples c04–c06. For each example, the top row shows weak pre-abstraction and the bottom row shows strong pre-abstraction. In the leftmost column, the top image is Sketch applied directly to the original image and the bottom image is the original photograph.