Supplementary of Spec2Slide: Agentic Image-Based Slide Generation with Iterative Feedback-Driven Optimization

Anonymous CVPR submission

Paper ID 21766

1. Image Disclaimer

All images in this work are synthetically generated by AI models and are used solely for non-commercial academic research purposes. They are not official assets of any third-party brand or rights holder. We do not intend to infringe any copyrights, trademarks, or other intellectual-property rights, and we claim no ownership over any recognizable brand-related elements that the images may resemble. Readers should not reuse these images for commercial or production use and are responsible for obtaining any necessary permissions for downstream applications.

2. Baselines

To facilitate reproducibility, the supplementary material includes all generated PPTX decks from the baseline systems (Gamma, Beautiful.ai, GenSpark, and AutoPresent), together with the full specifications of our dataset. For Gamma, we use a Plus subscription and invoke the "Create from notes or existing content" workflow, selecting the photorealistic option for image generation. For Beautiful.ai, we use a paid subscription, upload the slide content, and choose the Warm style, which most closely resembles a photorealistic look among the available options. For GenSpark, we use the Slides agent and pass the full slide content as input. For AutoPresent, we call the Gradio client API to generate slides and, whenever the programmatic slide generation produces a blank slide, we simply re-issue the API request.

Note: To reduce legal risk, we removed the AI-generated PPTX decks that depict Ford or Nike products from the baseline PPTX bundle in the supplementary materials, but we include the corresponding text specifications so that others can reproduce our results.

3. Evaluation prompt

We evaluate each generated slide with Gpt-4o using the exact prompt below. The model returns eight 1-5 integer ratings with brief explanations and an overall_score equal to the arithmetic mean (rounded to one decimal). Placeholders { original_spec }, {generated_prompt} and the image handle are filled at runtime.

```
You are an expert slide design and visual
     communication evaluator. Rate how
    well a GENERATED SLIDE IMAGE aligns
    with its ORIGINAL SLIDE SPEC and
    GENERATED SLIDE PROMPT.

You have access to the image content (
    assume vision). If an image is
    missing, set every rating to 1 with
    explanation "no image provided".

Provide 1-5 integer ratings (1=poor, 3=
    adequate, 5=excellent) and concise
    (<=2 sentence) explanations per
    criterion.

RUBRIC CRITERIA (all required):
1. theme_compliance
  Focus: Adherence to declared theme
      tokens (mood, typography class,
      stylistic constraints). Penalize
      off-theme fonts or visual motifs.
  5: Strongly on-theme; no contradictory
      elements.
  3-4: Mostly aligned; minor
      inconsistencies.
  1-2: Theme largely ignored or clashing.

2. layout_adherence
  Focus: Spatial arrangement vs prompt/
      spec (title position, body regions,
       imagery placement, hierarchy).
  5: Clear structural fidelity and
      hierarchy.
  3-4: Minor spacing/ordering drift.
```

CVPR #21766

```
    1-2: Major structural divergence /
        chaotic layout.

3. content_coverage
  Focus: Presence of required content
      regions (sections, bullet groups,
      imagery placeholders). Ignore
      literal text wording; measure
      structural presence.
  5: All required regions present.
  3-4: One minor region missing or merged
      .
  1-2: Multiple or critical regions
      absent.

4. image_quality
  Focus: Visual clarity (resolution/
      legibility), absence of severe
      artifacts, balanced negative space.
  5: Crisp, clean, professional quality.
  3-4: Minor artifacts or mild clutter.
  1-2: Noticeable artifacts, heavy
      clutter, or blurry.

5. illustration_style_alignment
  Focus: Illustration / imagery medium &
      stylistic traits match the
      specified illustration style (e.g.,
       Oil Painting, Flat Design). If no
      imagery expected AND none shown,
      rate 5.
  5: Style precisely matches descriptors.
  3-4: Generally correct; a few missing
      stylistic cues.
  1-2: Wrong medium, off-topic, or
      missing when required.

6. slide_quality
  Focus: Holistic professional polish (
      cohesion, balance, absence of
      distracting elements) beyond
      individual criteria.
  5: Production-ready; cohesive and
      polished.
  3-4: Usable with minor refinements.
  1-2: Rough / amateurish / inconsistent.

7. accessibility
  Focus: Readability & contrast;
      avoidance of text over busy
      backgrounds; sufficient spacing for
       scanning.
  5: High contrast, legible, good spacing
      .
  3-4: Minor readability or contrast
      issues.
  1-2: Poor contrast / overlapping / hard
       to read.

8. color_palette
  Focus: Fidelity to theme/prompt palette
       (primary, accent, neutrals)
      without clashing or unmotivated
      hues.
  5: Palette roles respected; cohesive
      harmony.
  3-4: Slight drift or extra mild accent.
  1-2: Off-theme, clashing, or
      inconsistent.

OUTPUT JSON ONLY (no prose, no markdown):
{
  "theme_compliance": {"rating": int, "
      explanation": ""},
  "layout_adherence": {"rating": int, "
      explanation": ""},
  "content_coverage": {"rating": int, "
      explanation": ""},
  "image_quality": {"rating": int, "
      explanation": ""},
  "illustration_style_alignment": {"
      rating": int, "explanation": ""},
  "slide_quality": {"rating": int, "
      explanation": ""},
  "accessibility": {"rating": int, "
      explanation": ""},
  "color_palette": {"rating": int, "
      explanation": ""},
  "overall_score": float
}
Overall_score = arithmetic mean of the
    eight ratings (round to one decimal
    if needed).

INPUT BLOCKS:
--- ORIGINAL_SPEC_START ---
{original_spec}
--- ORIGINAL_SPEC_END ---
--- GENERATED_PROMPT_START ---
{generated_prompt}
--- GENERATED_PROMPT_END ---
--- IMAGE_CONTEXT_START ---
(Provide any reference to or description
    / encoded handle of the generated
    image here.)
--- IMAGE_CONTEXT_END ---

Return ONLY the JSON.
```

4. Illustration Styles

Illustration styles used in our experiments.

1. **Anime:** Cel-shaded 2D illustration with clean lines and flat colors.
2. **Film Still:** Cinematic, hyperreal imagery with soft lighting and shallow depth of field.
3. **Neon Noir:** Neon-lit futuristic cityscapes with dense

CVPR
#21766

detail and atmospheric haze.

4. **Mini Model:** Stylized miniature 3D diorama with clean geometry and soft lighting.
5. **Loose Squiggles:** Playful black-and-white doodles with thick, energetic marker lines.
6. **Pure Vector:** Minimal vector graphics with simple shapes and flat solid colors.
7. **ShapeCraft 3D:** 3D geometric forms combining clean shapes with subtle textures.
8. **Axon 3D:** Isometric 3D scenes with rounded forms and soft, even lighting.
9. **Block 3D:** Angular low-poly 3D shapes with muted colors and simple shading.
10. **Inked Comic:** Monochrome manga inking with bold blacks and expressive texture.
11. **Concrete Forms:** Heavy concrete-like forms with strong geometry and deep shadows.
12. **Glow Stack:** Neon tubes, saturated glows, and reflections in dark environments.
13. **Oil Painting:** Thick impasto strokes with bold texture and high-contrast color.
14. **Hyperreal Lens:** Ultra-detailed portraits with natural color and shallow depth of field.
15. **Voxel Build:** Voxel-style 3D built from chunky cubes with game-like simplicity.
16. **8-Bit Grid:** Retro 8-bit pixel art with low resolution and limited palette.
17. **Spot-Color Print:** 2-3 color Risograph look with misregistration and halftone texture.
18. **Chaotic Lines:** Overlapping scribbled lines where density encodes form and shading.
19. **Dreamspace:** Hyperreal yet dreamlike scenes with abstract, gravity-defying forms.
20. **Outrun Glow:** 1980s-style neon gradients, horizon grids, and retro-futuristic geometry.
21. **Splash Wash:** Loose watercolor washes and splashes around a sharp focal point.
22. **Soft-Gradient 3D:** Rounded 3D forms with smooth gradients and playful, polished finishes.