VecGlypher: Unified Vector Glyph Generation with Language Models

Supplementary Material

1. Dataset Statistics (Cont')

Training data before filtering. Figure 1 visualizes the token-length distributions of the *training* corpora *before* applying the typography-specific filtering described in Sec. 3.3 of the main paper.

For each font, we compute:

- the **input token length**: the number of tokens in the serialized style description (for text-referenced samples), and
- the **output token length**: the number of tokens in the tokenized SVG `<path>` string for each glyph.

The top row of Fig. 1 shows the distributions for Google Fonts; the bottom row shows Envato.

Heavy-tailed SVG sequences. On both corpora, input style tokens are relatively concentrated around a narrow range (roughly one to two dozen tags per font, consistent with Fig. 2 in the main text).

In contrast, the *SVG path* token lengths are strongly long-tailed, especially for Envato: a small fraction of fonts contain glyphs whose serialized paths span tens of thousands of tokens. These extremely long sequences mostly correspond to malformed outlines, redundant contour duplication, or decorative symbols.

Such heavy tails are problematic for autoregressive training: they increase sequence entropy, cause unstable gradients, and exacerbate error accumulation during decoding. This motivates the length-based pruning strategy ("Length by pangram") described in Sec. 3.3 of the main paper.

Filtered test-set statistics. Figure 2 reports analogous statistics for the *testing* split, after all filtering steps and split-by-family partitioning. Compared to Fig. 1, the output token distributions are significantly better behaved:

- The vast majority of glyphs fall into a moderate length regime, enabling stable training and evaluation.
- Google Fonts remains more compact than Envato, consistent with its expert-curated nature and more regular outlines.

These plots confirm that the preprocessing pipeline successfully removes pathological fonts while retaining a broad diversity of typographic structure: after filtering we keep 2,497 Google Fonts and 39,497 Envato fonts, with 157,899 and 2,495,363 glyphs respectively. We exclude Envato fonts from testing because their tags are generally noisy and lack meaningful visual descriptions.

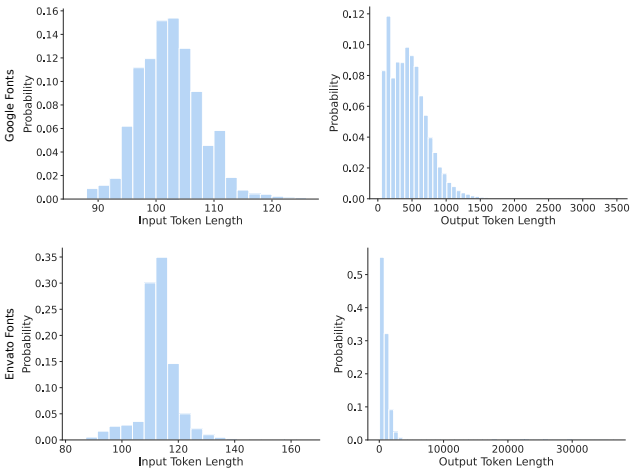

Figure 1. Dataset Statistics for training set before filtering.

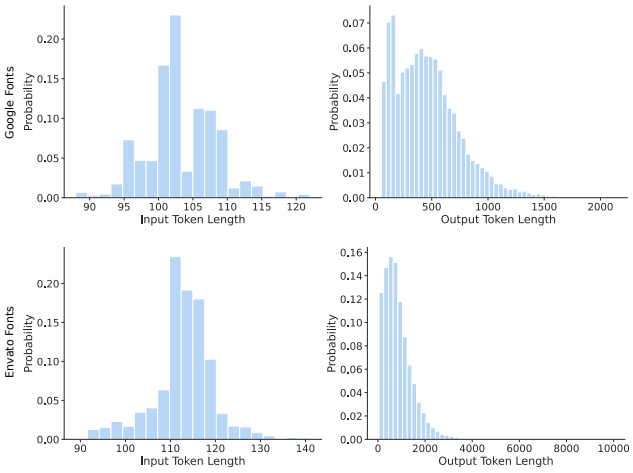

Figure 2. Dataset Statistics for testing set.

2. Prompt Templates and Samples

We use a strict prompting scheme to encourage syntactically valid SVG paths and to cleanly separate role instructions, font style descriptions, and target content. The main paper briefly mentions a "strict system prompt" for SVG-only outputs; here we describe the full templates, which are summarized in Table 1 of the supplementary material. The system prompt and the instruction prompts for both text- and image-referenced vector glyph generation. The "{{FONT STYLE}}" is the bag of style tags, and the "{{GLYPH CHARACTER}}" is a single given character in "0–9, a–z, A–Z".

System Prompt

You are a specialized vector glyph designer creating SVG path elements.
Critical requirements:
- Each glyph must be a complete, self-contained <path> element, in reading order of the given text.
- Terminate each <path> element with a newline character.
- Output ONLY valid SVG <path> elements.

Instruction Prompt for Text-Referenced Vector Glyph Generation

Font design requirements: {{FONT STYLE}}.
Text content: {{GLYPH CHARACTER}}.

Instruction Prompt for Image-Referenced Vector Glyph Generation

Font design requirements: faithfully match the provided reference images for style and metrics.
Text content: {{GLYPH CHARACTER}}.

Samples of instruction prompts for text-referenced vector glyph generation 1

Font design requirements: stiff, wordspace quality, superellipse sans, rounded sans-serif, 400 weight, sans-serif, normal style, display, neo grotesque sans-serif, drawing quality, cute.
Text content: V.

Samples of instruction prompts for text-referenced vector glyph generation 2

Font design requirements: wordspace quality, rounded sans-serif, sans-serif, competent, grotesque sans, 900 weight, italic style, calm.
Text content: b.

Samples of instruction prompts for text-referenced vector glyph generation 3

Font design requirements: fancy, 400 weight, vintage, sophisticated, handwritten script, brush, wordspace quality, handwriting, sincere, informal script, drawing quality, excited, normal style, cute, artistic.
Text content: 6.

3. Additional Metrics

The main paper evaluates VecGlypher and baselines with Relative OCR Accuracy (R-ACC), Chamfer Distance (CD), CLIP similarity, DINO similarity, and FID.

Here we introduce several additional metrics that capture complementary aspects of glyph quality and provide full results in Tables 3–6 of the supplementary.

All raster-based metrics are computed on 192x192 grayscale renderings, using the same rasterization pipeline as for the qualitative figures.

We further introduce:

- **FID (C).** In addition to FID computed in Inception space, we report **FID (C)**, where both real and generated glyphs are embedded with CLIP ViT-B/32 features before computing the Fréchet distance. This emphasizes semantic similarity between the stylized glyph images and is more sensitive to high-level appearance than to low-level pixel noise.
- **R-ACC (U)**: R-ACC in the main paper measures OCR accuracy normalized by the accuracy on ground-truth glyphs, allowing scores slightly above 100 because of OCR variability.
- **R-ACC (U)**: The OCR outputs are case-normalized: upper- and lowercase characters that share the same identity (e.g., "a" vs "A") are treated as correct. This disentangles shape-level recognition errors from case mismatches.
- **CD (T) and CD (ST)**: Our primary Chamfer Distance (CD) computes a symmetric distance between two point clouds sampled from the SVG outlines after normalizing each glyph to the $[-1, 1]$ box.
- **CD (T)**: we first align prediction to ground truth via Iterative Closest Point (ICP) optimizing only a 2D translation. This compensates for small global shifts, e.g., from imperfect baseline alignment.
- **CD (ST)**: we run ICP optimizing both translation and isotropic scale. This is suitable for italic or script fonts where rotation is ambiguous but overall scale may drift. We intentionally do *not* optimize rotation in either variant, since many italic fonts are skewed and rotational alignment could distort their intended slant.
- **L2**:**L2** is the mean squared error between rasterized predictions and ground truths, averaged over pixels and glyphs. It captures dense per-pixel discrepancies but is insensitive to human perception.
- **LPIPS**: We report Learned Perceptual Image Patch Similarity (LPIPS) using a standard VGG backbone. LPIPS measures perceptual distance between images by comparing deep feature activations, correlating better with human judgement than L2.
- **PSNR and SSIM**: We additionally report Peak Signal-to-Noise Ratio (PSNR) and Structural Similarity Index (SSIM), both computed on grayscale rasterizations. PSNR summarizes global reconstruction fidelity; SSIM emphasizes local luminance, contrast, and structural agreement.

4. Additional Baselines

To complement the proprietary LLMs evaluated in the main paper, we benchmark two classes of publicly available models on text-referenced glyph generation:

- **Open-weight multimodal LLMs**: Llama 3.3 70B Instruct, Gemma 3 27B IT, and Qwen 3 30B A3B Instruct (2507). These models are strong general-purpose multimodal assistants.
- **Vector-graphics LLM**: OmniSVG, an open-source model trained specifically for SVG generation and image vectorization. We skip LLM4SVG and StarVector since their text-to-SVG models are not publicly released.

Overall, we observe:

- **Extremely low recognizability.** R-ACC and R-ACC(U) remain close to zero for all open-weight LLMs and for OmniSVG, meaning that the OCR engine rarely recognizes the intended characters.
- **Poor geometry and image quality.** CD, CD(T), and CD(ST) are an order of magnitude higher than those of VecGlypher, while L2, LPIPS, and SSIM indicate severe geometric distortions or failure to render the glyph at all.
- **Limited benefit from SVG specialization.** OmniSVG, despite being trained for SVG icons and vectorization, produces particularly poor results for glyphs: a large fraction of outputs are invalid paths or generic shapes unrelated to the requested character or style.

These trends mirror the conclusions of Sec. 1 and Sec. 5 in the main paper: off-the-shelf LLMs and vector-graphics LLMs that perform well on icons or simple SVG drawings do not transfer to typography, which imposes stricter geometric, stylistic, and topological constraints.

5. Additional Qualitative Results

Please refer to the HTML pages for comprehensive results. They include both ablation studies and comparisons for text-referenced and image-referenced vector-glyph generation tasks.

Table 1. Text-referenced ablations on data and model size.

Data	Size	Repr.	R-ACC	CD	CLIP	DINO	FID	FID (C)	R-ACC (U)	CD (T)	CD (ST)	L2	LPIPS	PSNR	SSIM
Google	4B	Rel.	73.96	4.29	26.02	90.27	15.57	1.84	75.51	3.83	4.40	0.196	0.245	8.30	0.643
	4B	Abs.	66.66	3.75	26.04	89.92	14.69	2.19	68.68	3.36	3.85	0.205	0.251	8.14	0.631
	27B	Rel.	92.81	2.31	26.38	93.01	5.81	0.440	93.40	1.96	2.32	0.158	0.212	9.21	0.679
	27B	Abs.	94.91	1.98	26.43	93.43	3.96	0.250	95.30	1.69	2.00	0.152	0.204	9.50	0.686
Envato	27B	Rel.	93.99	3.89	26.43	89.79	30.68	1.63	94.86	3.35	3.90	0.211	0.265	7.46	0.621
	27B	Abs.	93.84	3.63	26.45	90.24	20.43	1.31	95.18	3.22	3.64	0.197	0.254	7.78	0.636
E + G	27B	Rel.	94.39	2.56	26.39	93.17	5.83	0.458	94.69	2.20	2.57	0.147	0.200	9.60	0.692
	27B	Abs.	95.59	2.14	26.45	93.66	4.86	0.289	96.27	1.84	2.14	0.141	0.194	9.81	0.698
E → G	27B	Rel.	99.88	1.71	26.52	94.07	3.57	0.142	99.90	1.44	1.71	0.141	0.194	9.74	0.697
	27B	Abs.	101.0	1.67	26.53	94.34	3.47	0.135	100.72	1.40	1.65	0.138	0.191	9.88	0.700

Table 2. Image-referenced ablations on data and model size.

Data	Size	Repr.	R-ACC	CD	CLIP	DINO	FID	FID (C)	R-ACC (U)	CD (T)	CD (ST)	L2	LPIPS	PSNR	SSIM
Google	4B	Rel.	83.48	2.36	25.90	93.26	9.38	1.08	84.66	1.97	2.45	0.155	0.206	9.60	0.684
	4B	Abs.	81.94	2.11	25.89	93.11	7.94	1.07	83.53	1.77	2.19	0.156	0.205	9.58	0.682
	27B	Rel.	94.42	1.53	26.03	94.88	4.07	0.293	95.30	1.24	1.50	0.127	0.179	10.65	0.716
	27B	Abs.	96.46	1.41	26.04	95.15	2.84	0.147	97.03	1.12	1.34	0.121	0.172	10.79	0.722
E-G (I)	27B	Rel.	98.90	1.17	26.06	95.69	2.51	0.116	99.26	0.910	1.09	0.109	0.159	11.22	0.736
	27B	Abs.	97.88	1.16	26.06	95.84	2.55	0.101	98.43	0.910	1.09	0.107	0.156	11.36	0.740
E-G (T+I)	27B	Rel.	99.08	1.21	26.07	95.65	2.48	0.120	99.44	0.950	1.13	0.112	0.162	11.38	0.734
	27B	Abs.	99.12	1.18	26.07	95.82	2.32	0.095	99.82	0.930	1.10	0.110	0.158	11.35	0.736

Table 3. Text-referenced comparisons with general LLMs.

Model	R-ACC	CD	CLIP	DINO	FID	FID (C)	R-ACC (U)	CD (T)	CD (ST)	L2	LPIPS	PSNR	SSIM
GPT-5 mini	4.17	10.70	24.75	79.65	63.86	13.78	4.96	10.49	10.74	0.382	0.412	4.45	0.432
Gemini-2.5 Flash	4.89	13.78	25.26	78.62	86.03	12.77	7.15	13.38	13.78	0.364	0.379	4.85	0.480
Claude Haiku 4.5	32.38	7.60	25.61	84.69	35.07	5.85	38.80	7.17	7.60	0.289	0.332	5.96	0.545
GPT-5	43.98	6.12	25.95	86.92	29.00	3.71	48.03	5.26	6.26	0.301	0.346	5.68	0.521
Gemini 2.5 Pro	24.04	8.22	25.57	83.77	47.82	7.39	27.06	7.72	8.22	0.319	0.347	5.48	0.518
Claude Sonnet 4.5	46.65	5.28	25.99	88.31	19.59	2.81	52.99	4.58	5.34	0.253	0.305	6.62	0.580
VecGlypher 27B T,I,R	99.62	1.77	26.53	93.97	3.50	0.139	99.58	1.51	1.76	0.143	0.197	9.64	0.695
VecGlypher 27B T,I,A	100.5	1.72	26.53	94.22	3.46	0.134	100.34	1.44	1.68	0.140	0.193	9.83	0.698
VecGlypher 70B T,R	100.1	1.70	26.53	94.10	3.45	0.140	100.57	1.43	1.67	0.140	0.193	9.80	0.699
VecGlypher 70B T,A	100.4	1.68	26.54	94.28	3.34	0.136	100.71	1.40	1.66	0.139	0.192	9.85	0.699

Table 4. Image-referenced comparisons with vector-font baselines.

Model	R-ACC	CD	CLIP	DINO	FID	FID (C)	R-ACC (U)	CD (T)	CD (ST)	L2	LPIPS	PSNR	SSIM
DeepVecFont-v2	37.86	14.58	24.81	79.41	115.5	16.45	49.29	14.58	14.58	0.243	0.320	6.70	0.599
DualVector	49.20	16.45	25.07	79.57	105.5	14.73	65.59	16.44	16.44	0.190	0.293	7.99	0.653
VecGlypher 27B T,I,R	99.08	1.21	26.07	95.65	2.48	0.120	99.44	0.950	1.13	0.112	0.162	11.38	0.734
VecGlypher 27B T,I,R	99.12	1.18	26.07	95.82	2.32	0.095	99.82	0.930	1.10	0.110	0.158	11.35	0.736

Table 5. Performance of the open weight LLMs and the vector graphic LLM.

Model	R-ACC	CD	CLIP	DINO	FID	FID (C)	R-ACC (U)	CD (T)	CD (ST)	L2	LPIPS	PSNR	SSIM
Llama3.3 70B Instruct	0.08	38.28	24.84	75.85	143.73	21.49	0.08	38.26	38.30	0.374	0.388	4.64	0.480
Gemma3 27B IT	0.12	17.90	24.78	74.26	162.89	23.24	0.11	17.44	17.90	0.407	0.397	4.28	0.450
Qwen3 30B A3B Instruct 2507	0.39	26.15	24.66	74.03	148.78	22.47	0.55	26.07	26.15	0.418	0.422	4.219	0.427
OmniSVG	0.01	63.70	23.26	69.35	229.38	35.75	0.01	63.69	63.69	0.496	0.452	3.50	0.356

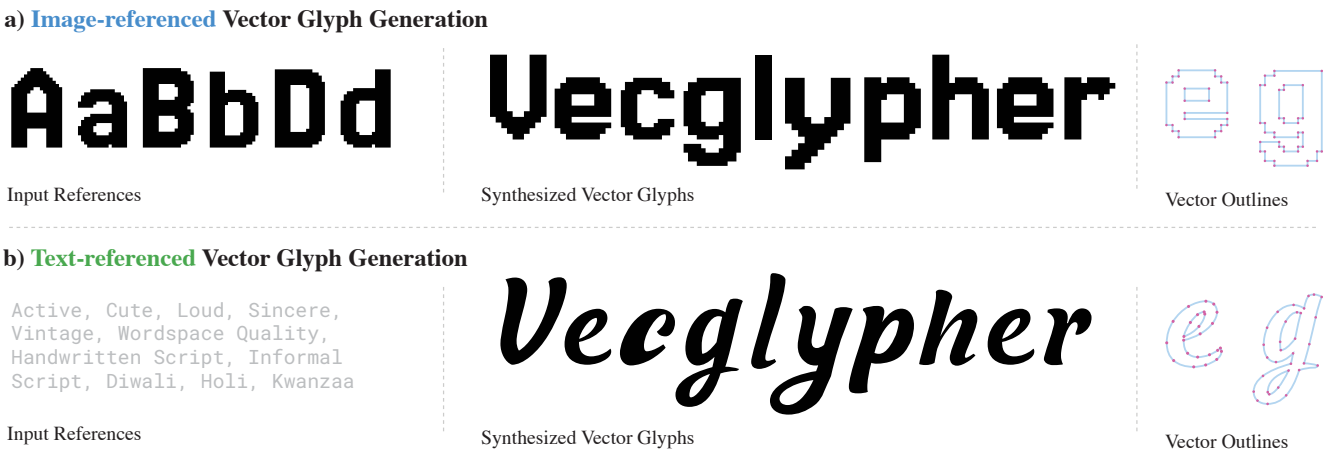

Figure 3. **VecGlypher** generates high-fidelity vector glyphs directly as editable SVG outlines under two types of conditioning: (a) **image-referenced** generation, where a handful of exemplar glyph images specify the style and the model synthesizes new glyphs in the same visual form; and (b) **text-referenced** generation, where a natural-language prompt drives the synthesis without requiring exemplars. The figure shows the synthesized wordmark and sample vector outlines, highlighting one-pass generation of clean, controllable contours for typography workflows.