Distilling Quasi-Conformal Mapping: A Generalizable and Efficient Solution for Wide-Angle Correction

Supplementary Material

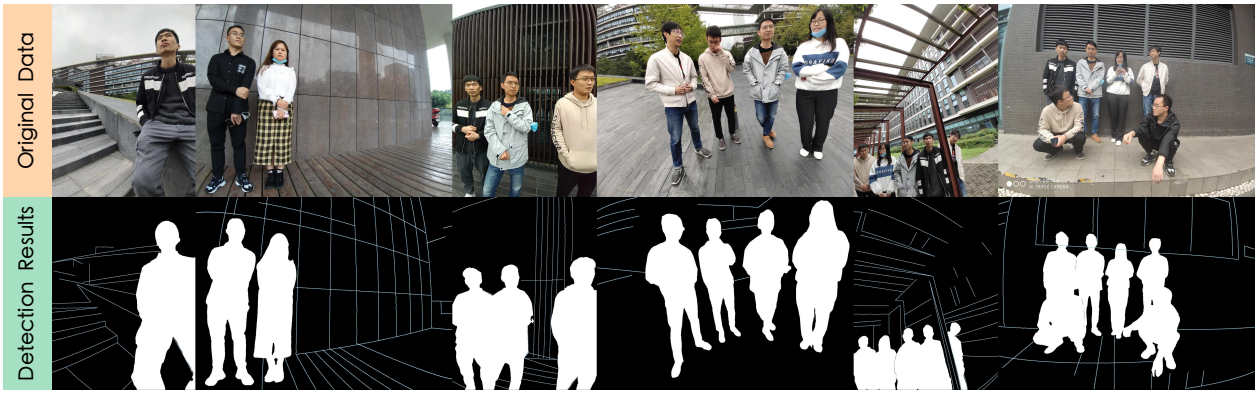

Figure S1. Illustration of line structure and human body region detection.

A. Implementation Details of Line and Mask Detection

Given a distorted image I_0, first, we use the YOLOv13 model [4] to segment human bounding boxes and then employ the SAM2 model [5] to generate masks for human body regions. We apply morphological image processing (*e.g.*, erosion and dilation) to remove noise and obtain accurate binary masks. Next, we white out the masked regions in the original images before performing line detection with the L-CNN model [8], thereby preventing the detection of lines on human bodies. Selected results of line structures and human body region detection are shown in Figure S1.

Some wide-angle images suffer from curved edge lines due to device lens distortion. We adopt the formula from [1] for rectification:

$$r_u = \frac{r_d}{1 + k_1 r_d^2 + k_2 r_d^4 + \cdots}, \tag{1}$$

where r_d denotes the distance from a pixel to the image center, r_u is the distance from the corresponding pixel to the image center in the rectified image, and $k_1, k_2, \ldots$ are lens distortion coefficients. Using this formula, we obtain the mapping f_{ori2line} and the linear perspective projection image $I_{\text{w/o lens distortion}}$.

We then feed $I_{\text{w/o lens distortion}}$ into the L-CNN model for line detection. To restore the detection results to the original image coordinate system, we represent each detected line segment as a discrete set of points through equidistant sampling. For a line segment with endpoints $\mathbf{p}_1$ and $\mathbf{p}_2$ in the rectified domain, the sampled points $\mathbf{q}_j$ are calculated as:

$$\mathbf{q}_j = \mathbf{p}_1 + \frac{j}{N}(\mathbf{p}_2 - \mathbf{p}_1), \quad j = 0, 1, \ldots, N, \tag{2}$$

where the sampling density N is proportional to the Euclidean length of the segment. Finally, these discrete points are restored to the original image coordinate system via the correspondence: $(h, w) \in I_{\text{w/o lens distortion}} \leftrightarrow (f_{\text{ori2line},x}(h, w), f_{\text{ori2line},y}(h, w)) \in I_0$.

B. Details of Evaluation Metrics

LineAcc [6] evaluates the straightness of corrected structural lines:

$$\Delta_k = \left| \frac{y_k - y_{k-1}}{x_k - x_{k-1} + \varepsilon} - S_g \right|, \quad S_g = \frac{y_n - y_0}{x_n - x_0 + \varepsilon}, \tag{3}$$

$$\text{LineAcc} = 1 - \frac{1}{n} \sum_{k=1}^{n} \frac{\Delta_k - \Delta_{\min}}{\Delta_{\max} - \Delta_{\min} + \varepsilon}, \tag{4}$$

where Δ_k is the slope difference between the k-th segment and S_g, and ε is used to avoid division by zero.

FaceAcc [6] evaluates the consistency of corrected facial shapes:

$$\cos_i = \frac{\langle \mathbf{v}_i, \mathbf{v}_i^{\text{gt}} \rangle}{\|\mathbf{v}_i\| \, \|\mathbf{v}_i^{\text{gt}}\| + \varepsilon}, \quad \begin{aligned} \mathbf{v}_i &= \mathbf{p}_i - \mathbf{p}_{\text{nose\_tip}}, \\ \mathbf{v}_i^{\text{gt}} &= \mathbf{p}_i^{\text{gt}} - \mathbf{p}_{\text{nose\_tip}}^{\text{gt}}, \end{aligned} \tag{5}$$

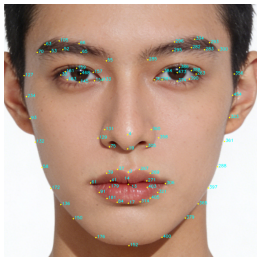

Figure S2. 81 facial landmarks. The portrait is a synthetic face generated by Doubao AI [1].

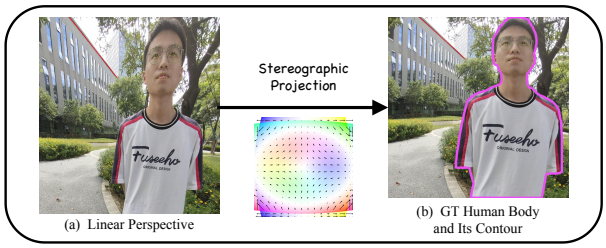

Figure S3. Illustration of the ground-truth human body and its contour, where each ground truth is individually optimized by leveraging stereographic projection during the test set annotation.

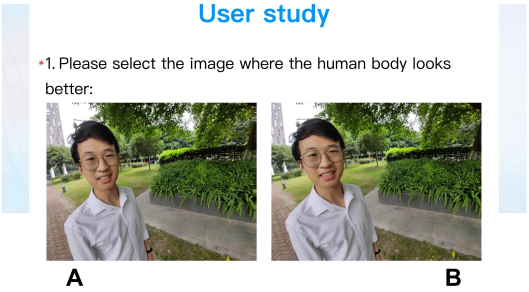

Figure S4. Sample question from our user study.

$$\text{FaceAcc} = \frac{1}{n-1}\sum_{i=1}^{n}\cos_i, \tag{6}$$

where $\mathbf{p}_i$ and $\mathbf{p}_i^{\text{gt}}$ are the predicted and ground-truth facial landmarks, respectively. For the unlabeled data (*e.g.*, Zhu's dataset [9]), we employ the MediaPipe Face Landmarker [2] to obtain the landmarks. To ensure alignment with Tan's method [6], we select 81 specific facial landmarks, as visualized in Figure S2. Their corresponding indices include 0, 4, 13, 14, 17, 19, 33, 37, 39, 41, 52, 53, 55, 58, 61, 63, 65, 66, 70, 75, 84, 91, 93, 105, 127, 129, 131, 132, 133, 136, 145, 150, 152, 154, 157, 159, 161, 163, 172, 176, 179, 181, 234, 263, 267, 269, 271, 282, 283, 285, 288, 293, 295, 296, 300, 305, 306, 314, 321, 323, 334, 356, 358, 360, 361, 362, 365, 374, 379, 381, 384, 386, 388, 390, 397, 400, 403, 405, 454, 468, and 473.

BodyAcc (Ours) evaluates the consistency of corrected human body regions. The ground-truth contour required in

[1] https://www.doubao.com

[2] https://github.com/google-ai-edge/mediapipe

Algorithm S1: Calculation of BodyAcc

Input: Image I, ground-truth contours $\mathcal{C}^{gt} = \{C_j^{gt}\}$, $n_{sampled} \gg 1$.
Output: Weighted human body similarity score $\mathcal{S}_{final} \times 100\%$.

Stage 1: Deep human body detection & segmentation.
$\mathcal{B} \leftarrow$ Human bounding boxes on I are detected by applying the YOLOvX model.
$\mathcal{M} \leftarrow \emptyset$. **for** B_i *in* $\mathcal{B}$ **do**
  $\mathcal{M} \leftarrow \mathcal{M} \cup$ A mask for B_i is generated by applying the SAM2 model.

Stage 2: Contour extraction, matching, & resampling.
$\mathcal{C} \leftarrow \emptyset$. **for** M_i *in* $\mathcal{M}$ **do**
  $\mathcal{C} \leftarrow \mathcal{C} \cup \{\texttt{findContours}\,(M_i)\}$.
Distance matrix $D \leftarrow [\ ]$. **for** C_i *in* $\mathcal{C}$ **do**
  for C_j^{gt} *in* $\mathcal{C}^{gt}$ **do**
    $D_{i,j} \leftarrow \left\|\texttt{mean}\,(C_i) - \texttt{mean}\,(C_j^{gt})\right\|_2$.
$\sigma \leftarrow$ The one-to-one assignment $(i \mapsto j)$ is obtained from D by applying the Hungarian algorithm.
Resampled contours $\mathcal{P}, \mathcal{Q} \leftarrow \emptyset, \emptyset$. Areas $\mathcal{A}^{Q} \leftarrow \emptyset$.
for $i = 1$ **to** $|\mathcal{C}^{gt}|$ **do**
  $A, A^{gt} \leftarrow \texttt{contourArea}(C_i)\,, \texttt{contourArea}\left(C_{\sigma(i)}^{gt}\right)$.
  $\mathcal{A}^{Q} \leftarrow \mathcal{A}^{Q} \cup \{A^{gt}\}$.
  $\text{index}, \text{index}^{gt} \leftarrow \emptyset, \emptyset$. **for** $k = 1$ **to** $n_{sampled}$ **do**
    $\text{index} \leftarrow \text{index} \cup \{\texttt{round}\,(1 + (k-1)\,(|C_i| - 1)\,/(n_{sampled} - 1))\}$,
    $\text{index}^{gt} \leftarrow \text{index}^{gt} \cup \left\{\texttt{round}\left(1 + (k-1)\left(\left|C_{\sigma(i)}^{gt}\right| - 1\right)/(n_{sampled} - 1)\right)\right\}$.
  $\mathcal{P} \leftarrow \mathcal{P} \cup \left\{\sqrt{\frac{A^{gt}}{A}}\,(C_i[\text{index}] - \texttt{mean}\,(C_i))\right\}$,
  $\mathcal{Q} \leftarrow \mathcal{Q} \cup \left\{C_{\sigma(i)}^{gt}\,[\text{index}^{gt}] - \texttt{mean}\left(C_{\sigma(i)}^{gt}\right)\right\}$.

Stage 3: Human body similarity computation.
Set of similarity scores $\mathcal{S} \leftarrow \emptyset$. **for** (P_k, Q_k) *in* $(\mathcal{P}, \mathcal{Q})$ **do**
  Fréchet distance $d_F(P_k, Q_k) \leftarrow \min_{\pi \in \Pi(m,n)} \max_{(i,j)\in\pi} \|p_i - q_j\|_2$, where $\Pi(m,n)$ denotes the set of all monotone paths from $(1,1)$ to $(n_{sampled}, n_{sampled})$, and $p_i, q_j \in P_k, Q_k$.
  $L, L^{gt} \leftarrow \texttt{arcLength}(P_k), \texttt{arcLength}(Q_k)$.
  Convert distance to similarity $s \leftarrow \max\left\{0, 1 - \frac{d_F(P_k, Q_k)}{\sqrt{L L^{gt}}/\sqrt{2}}\right\}$.
  $\mathcal{S} \leftarrow \mathcal{S} \cup \{s\}$.
$\mathcal{S}_{final} \leftarrow \sum_k \frac{\mathcal{A}_k^{Q}\mathcal{S}_k}{\sum_l \mathcal{A}_l^{Q}}$.

Algorithm S1 is annotated as illustrated in Figure S3. Furthermore, we conducted a **user study** with 52 participants and collected data on the proportion of choices consistent with our proposed BodyAcc. A Wilcoxon signed-rank test yielded $\texttt{W=1201}$ and $\texttt{p=1.34e-06} \ll \texttt{0.01}$, allowing us to accept the alternative hypothesis that perceptual naturalness aligns significantly with BodyAcc at the 99% confidence level.

The 204 annotated images from Zhu's dataset are available at https://drive.google.com/file/d/1gm0qjOgNK1CJCtPlmGlj0pGAbsX8UAbp/view?usp=sharing.

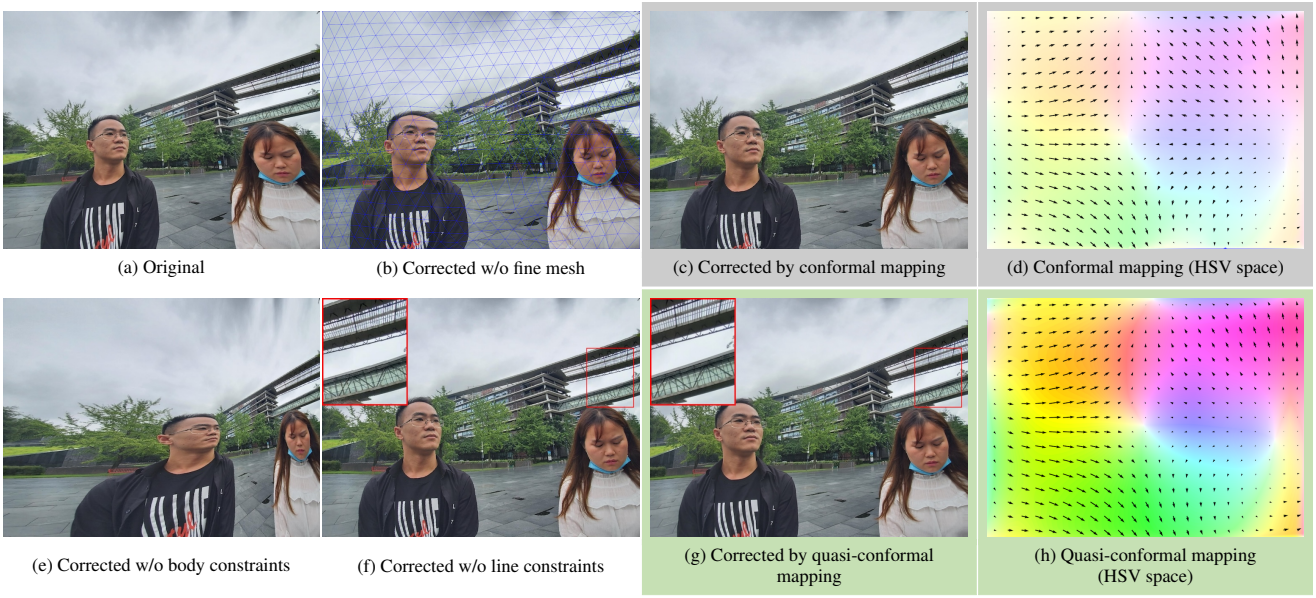

Figure S5. Ablating settings of quasi-conformal mapping. (a) The original wide-angle distorted image. (b) Correction with a coarse mesh (450 vertices). (c) The conformal mapping method. (d) Pixel-wise flow visualization of the conformal mapping. (e) Correction without human body constraints. (f) Correction without line structure constraints. (g) Our quasi-conformal mapping method with a fine mesh (5,000 vertices) and all constraints, yielding distortion-free results. (h) Pixel-wise flow visualization of the quasi-conformal mapping.

Table S1. Comparison with conformal mapping on Tan's dataset.

Metrics	Conformal	Quasi-conformal	Distilling Conformal	Distilling QC
LineAcc ↑	70.229	**70.431**	69.707	<u>70.294</u>
FaceAcc ↑	97.938	98.438	<u>98.527</u>	**98.607**
BodyAcc ↑	97.609	<u>97.780</u>	97.426	**97.868**

Table S2. Latency comparison of different mapping methods.

	Conformal (5,000 vertices)	Quasi-conformal	
		Coarser mesh (450 vertices)	Fine mesh (5,000 vertices)
Latency (s)	25.03	14.57	26.33

(a) Pre-reconstruction ($\sigma_{\text{top 4}} = 92.2\%$)
(b) Post-reconstruction ($\sigma_{\text{top 4}} = 99.9\%$)

Figure S6. Comparison of results before and after low-rank prior reconstruction. Here, $\sigma_{\text{top 4}}$ denotes the total energy of the correction flow captured by the top 4 singular values. (a) The initial correction result and confidence map. Low-confidence regions (blue masks) are magnified for a better view. (b) The post-reconstruction result.

C. Ablation Studies

C.1. Settings of Quasi-conformal Mapping

We evaluate the impact of mesh resolution, constraint conditions, and mapping paradigms on the quasi-conformal mapping method. The qualitative ablation results are illustrated in Figure S5. Table S1 and S2 also confirm that purely geometric quasi-conformal mapping outperforms conformal mapping at almost no additional computational cost, and distillation further improves performance, yielding the best overall results.

C.2. Backbone Architecture

As shown in Table S3, given comparable parameter counts, the Mamba backbone achieves the lowest FLOPs and latency while maintaining strong performance. This is due to the linear $\mathcal{O}(T)$ complexity of the Mamba architecture [2]. Additionally, the 16-layer Mamba backbone, which is employed in our model, achieves a balance between efficiency and performance.

C.3. Output Head Designs

Table S4 shows the effectiveness of the proposed output head. Specifically, the soft-argmin regression contributes to human body correction, while the low-rank prior ensures global consistency and recovers low-confidence regions (Figure S6). These components significantly enhance LineAcc, FaceAcc, and generalization on Zhu's dataset.

Table S3. Comparison of different backbone architectures on Tan's dataset.

Architectures	**#Param.** (M)	**FLOPs** (G)	LineAcc ↑	FaceAcc ↑	BodyAcc ↑	Latency ↓
CNN [3]	22.1	21.58	70.082	97.446	97.756	0.72 s
Transformer [7]	22.1	73.12	70.144	**98.623**	97.730	1.13 s
Mamba$_{\times 2}$	7.3	9.55	69.859	98.596	97.734	**0.46** s
Mamba$_{\times 4}$	9.4	10.47	69.957	98.567	97.788	0.49 s
Mamba$_{\times 8}$	13.7	12.32	69.880	98.600	97.783	0.57 s
Mamba$_{\times 16}$	22.1	16.02	**70.294**	98.607	**97.868**	0.67 s

Table S4. Quantitative ablation of the QDWC-Net output head. LineAcc ↑, FaceAcc ↑, BodyAcc ↑, Latency ↓. The numbers in parentheses indicate the sample count for each benchmark.

Architectures	**#Param.**	**FLOPs**	**Testing (129)**			**Generalization (204)**			
			LineAcc	FaceAcc	BodyAcc	LineAcc	FaceAcc	BodyAcc	Latency
Mamba$_{\times 16}$ backbone	21.9 M	15.88 G	69.992	98.578	97.767	87.461	98.744	92.356	**0.65** s
+ Soft-argmin w/o lowrank	22.1 M	16.02 G	70.294	98.607	**97.868**	87.442	98.937	92.339	0.67 s
+ Soft-argmin w/ lowrank	22.1 M	16.02 G	**70.299**	**98.619**	97.843	**88.627**	**98.972**	**92.499**	0.81 s

Table S5. Quantitative ablation of loss functions on Tan's dataset.

L_{flow}	L_{img}	L_{ssim}	LineAcc ↑	FaceAcc ↑	BodyAcc ↑
✔	✘	✘	70.119	98.562	97.763
✘	✔	✘	70.022	98.601	97.820
✔	✔	✘	70.100	98.612	97.666
✔	✔	✔	**70.299**	**98.619**	**97.843**

(a) Original BodyAcc: 96.418 (b) Teacher BodyAcc: 96.637 (c) Student BodyAcc: 97.198

Figure S7. Noisy label robustness: the teacher shows residual distortions; distillation yields a more robust student (Tan's dataset).

C.4. Loss Functions

Table S5 shows that the combination of L_{flow}, L_{img}, and L_{ssim} yields the best performance.

D. More Comparative Results

Figure S7 illustrates that the teacher is sensitive to line and mask detection errors: even with current SOTA detectors, missed or incorrect detections can produce residual distortions. However, by distilling from many such cases, the student is more robust to these noisy labels. Additional qualitative results are provided in Figure S8.

References

[1] Jinlong Fan, Jing Zhang, Stephen J. Maybank, and Dacheng Tao. Wide-angle image rectification: a survey. *International Journal of Computer Vision (IJCV)*, 130(3):747–776, 2022. 1

[2] Albert Gu and Tri Dao. Mamba: Linear-time sequence modeling with selective state spaces. In *First Conference on Language Modeling*, 2024. 3

[3] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 770–778, 2016. 4

[4] Mengqi Lei, Siqi Li, Yihong Wu, Han Hu, You Zhou, Xinhu Zheng, Guiguang Ding, Shaoyi Du, Zongze Wu, and Yue Gao. Yolov13: real-time object detection with hypergraph-enhanced adaptive visual perception. *arXiv preprint arXiv:2506.17733*, 2025. 1

[5] Nikhila Ravi, Valentin Gabeur, Yuan-Ting Hu, Ronghang Hu, Chaitanya Ryali, Tengyu Ma, Haitham Khedr, Roman Rädle, Chloe Rolland, Laura Gustafson, Eric Mintun, Junting Pan, Kalyan Vasudev Alwala, Nicolas Carion, Chao-Yuan Wu, Ross Girshick, Piotr Dollár, and Christoph Feichtenhofer. Sam 2: segment anything in images and videos. In *The 13th International Conference on Learning Representations (ICLR)*, 2025. 1

[6] Jing Tan, Shan Zhao, Pengfei Xiong, Jiangyu Liu, Haoqiang Fan, and Shuaicheng Liu. Practical wide-angle portraits correction with deep structured models. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 3498–3506, 2021. 1, 2

[7] Ashish Vaswani, Noam Shazeer, et al. Attention is all you need. In *Advances in Neural Information Processing Systems*, 2017. 4

[8] Yichao Zhou, Haozhi Qi, and Yi Ma. End-to-end wireframe parsing. In *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*, pages 962–971, 2019. 1

[9] Fushun Zhu, Shan Zhao, Peng Wang, Hao Wang, Hua Yan, and Shuaicheng Liu. Semi-supervised wide-angle portraits correction by multi-scale transformer. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 19689–19698, 2022. 2

Figure S8. Qualitative comparison of the proposed QDWC-Net with existing methods.