Layout-Aware Single-Image Document Flattening (Supplementary Material)

PU LI, MAIS, Institute of Automation, CAS and School of Artificial Intelligence, UCAS, China
WEIZE QUAN, MAIS, Institute of Automation, CAS and School of Artificial Intelligence, UCAS, China
JIANWEI GUO*, MAIS, Institute of Automation, CAS and School of Artificial Intelligence, UCAS, China
DONG-MING YAN, MAIS, Institute of Automation, CAS and School of Artificial Intelligence, UCAS, China

In this material, we provide more results as a supplement to our main paper. In Section A, we detail the relevant settings of the mesh simulation and the information for generating meshes, while in Section B, we perform ablation experiments on the network backbone.

A MESH SIMULATION PARAMETERS AND TIMES

Our simulation step does not reach a static equilibrium. We sample the simulated results during the mesh deformation under an external force. In the process of generating simulated deformation paper, using different numbers of added edges will generate meshes with different numbers of vertices and faces, and the strength of the force field should change accordingly to ensure realistic deformation. Therefore, the simulation time is related to the size of the mesh, *i.e.,* the larger the number of edges, the longer the simulation takes to generate a mesh. We list the range of simulation time corresponding to different numbers of edges in Tab. 1 and the related parameters of Blender in Tab. 2. It can be seen that simulation is more efficient than manual scanning and still produces realistic rendering results.

B ABLATION OF BACKBONES

We provide ablation studies on different network backbones. Specifically, we replace the original ViT-Hybrid backbone with ViT-Base, ViT-Large, and ResNet50, where all ViT backbones use 16×16 input patch sizes. We list the experimental results in Tab. 3, and the ViT-Hybrid backbone achieves the best results.

*Corresponding author: Jianwei Guo (jianwei.guo@nlpr.ia.ac.cn)

Authors' addresses: Pu Li, MAIS, Institute of Automation, CAS and School of Artificial Intelligence, UCAS, China; Weize Quan, MAIS, Institute of Automation, CAS and School of Artificial Intelligence, UCAS, China; Jianwei Guo, jianwei.guo@nlpr.ia.ac.cn, MAIS, Institute of Automation, CAS and School of Artificial Intelligence, UCAS, China; Dong-Ming Yan, MAIS, Institute of Automation, CAS and School of Artificial Intelligence, UCAS, China.

Table 1. Mesh generation details: the square brackets represent the numerical range, and strength denotes the turbulent force field parameter. Time signifies the required creation duration for a single mesh.

Edges Added	Vertices	Faces	Strength	Time (s)
[5,9]	[12,47]	[6,36]	40	0.031
[10,14]	[29,97]	[17,81]	30	0.113
[15,19]	[70,177]	[53,156]	20	0.272
[20,24]	[128,270]	[106,244]	15	0.607
[25,29]	[197,367]	[167,337]	15	0.809
[30,34]	[287,494]	[255,459]	10	1.219
[35,39]	[381,645]	[345,604]	10	1.547
[40,44]	[450,822]	[409,781]	10	1.778
[45,49]	[636,999]	[592,951]	10	2.339

Table 2. Parameter settings for Blender when generating meshes.

Object	Parameter Name	Setting
Mesh	Stiff Quads	TRUE
	Goal	TRUE
	Mass	1kg
	Bend	1
	Damp	0.5
	Simulation Speed	1
Turbulance	Shape	Point
	Noise Amount	10
	Strength	[10,40]
Scene	End Frame	10

Table 3. Ablation study on the backbones. Bold numbers are the best results, and the underlined setting means our final adoption.

Setting		(a)	(b)	(c)	(d)
Backbone	ResNet50	✓			
	ViT-Base		✓		
	ViT-Large			✓	
	ViT-Hybrid				✓
DocUNet Benchmark	**MS-SSIM**↑	0.500	0.485	0.507	**0.526**
	LD↓	8.028	7.961	7.701	**6.719**
	ED↓	505.1	666.8	542.5	**391.9**
	CER↓	0.188	0.214	0.183	**0.153**
Our Evaluation Dataset	**MPD**↓	4.044	4.044	3.811	**3.496**
	MS-SSIM↑	0.519	0.621	0.634	**0.677**
	LD↓	5.081	4.191	3.894	**3.643**
	ED↓	502.9	501.4	473.5	**365.4**
	CER↓	0.187	0.194	0.185	**0.135**