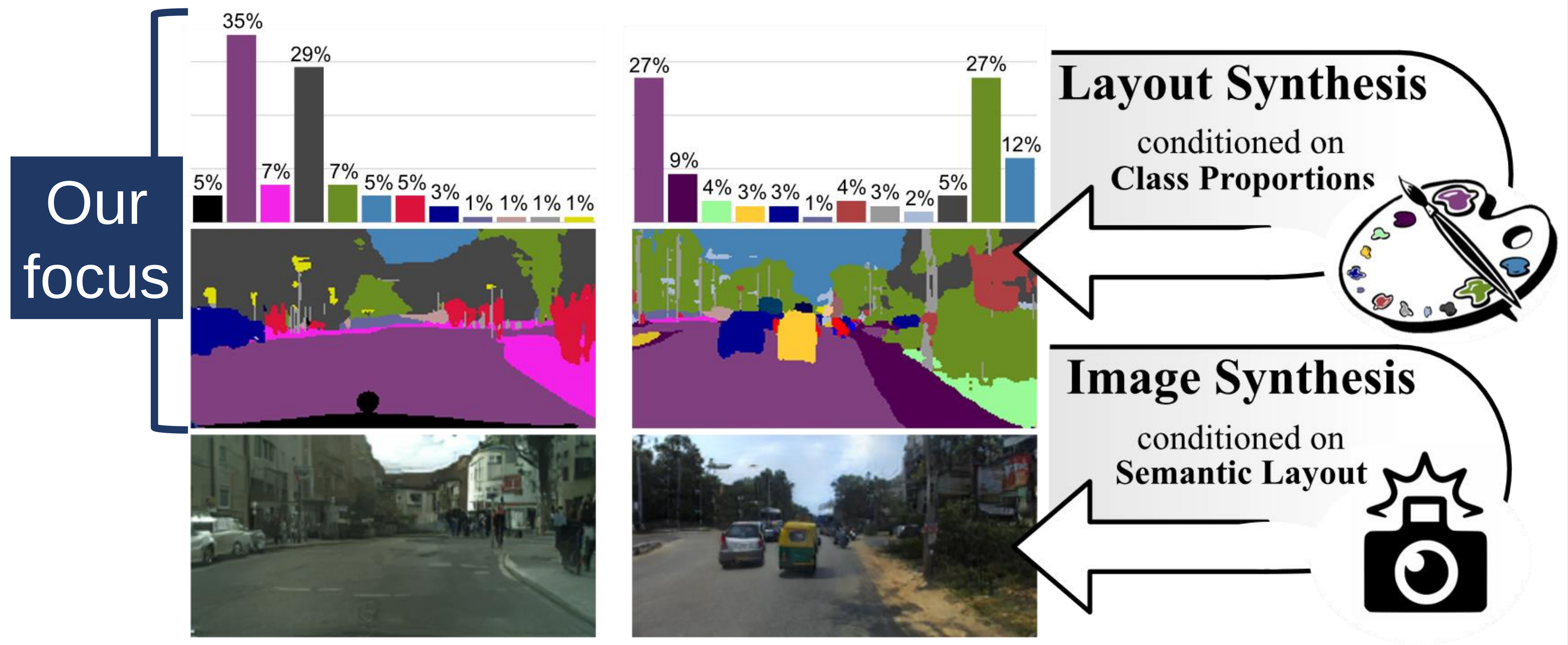
CVPR VIRTUAL JUNE 19-25

Semantic Palette: Guiding Scene Generation with Class Proportions

Guillaume Le Moing[1,†], Tuan-Hung Vu[2], Himalaya Jain[2], Patrick Pérez[2], Matthieu Cord[2]

[1]Inria, École normale supérieure, CNRS, PSL Research University, France [2]Valeo.ai [†]work performed during an internship at Valeo

Introduction



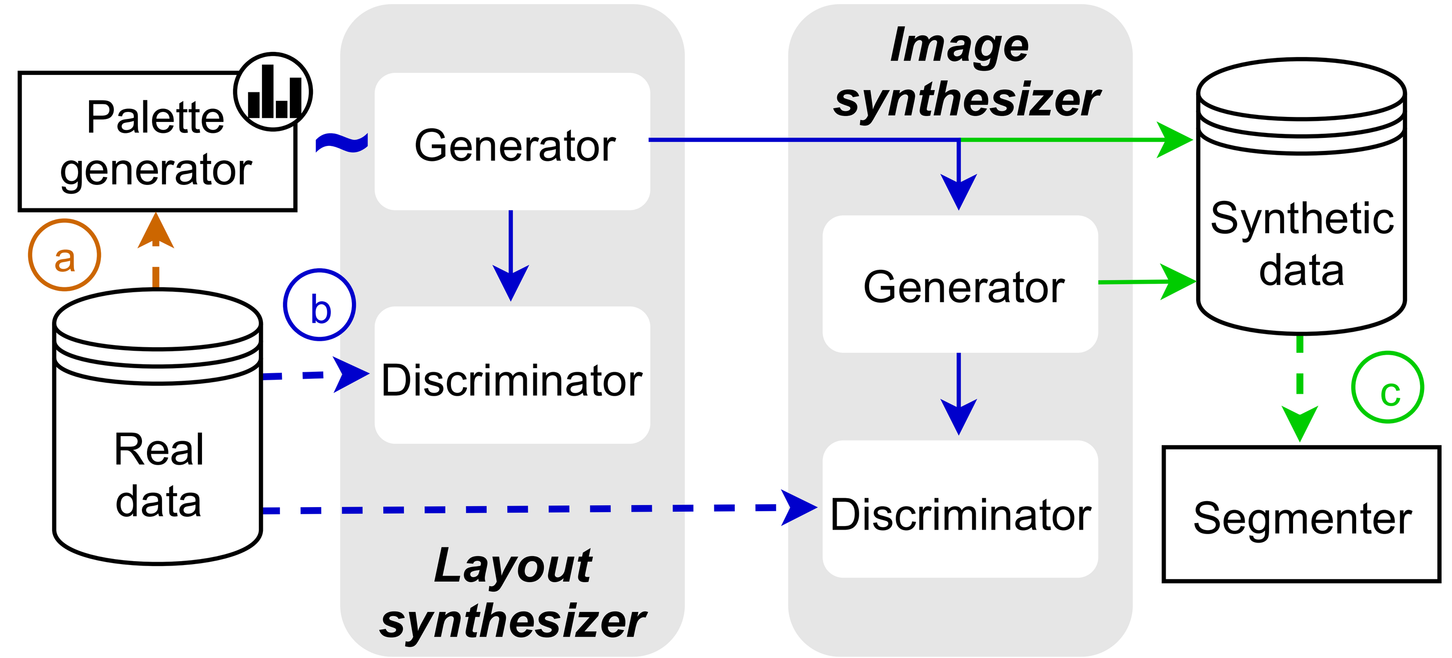
Towards controllable quality scene generation, we propose:

1. the task of layout synthesis conditioned on class proportions,
2. the combination of a layout generator and an image generator to produce images of greater practical use and quality,
3. the extension of our framework to partial editing of scenes.

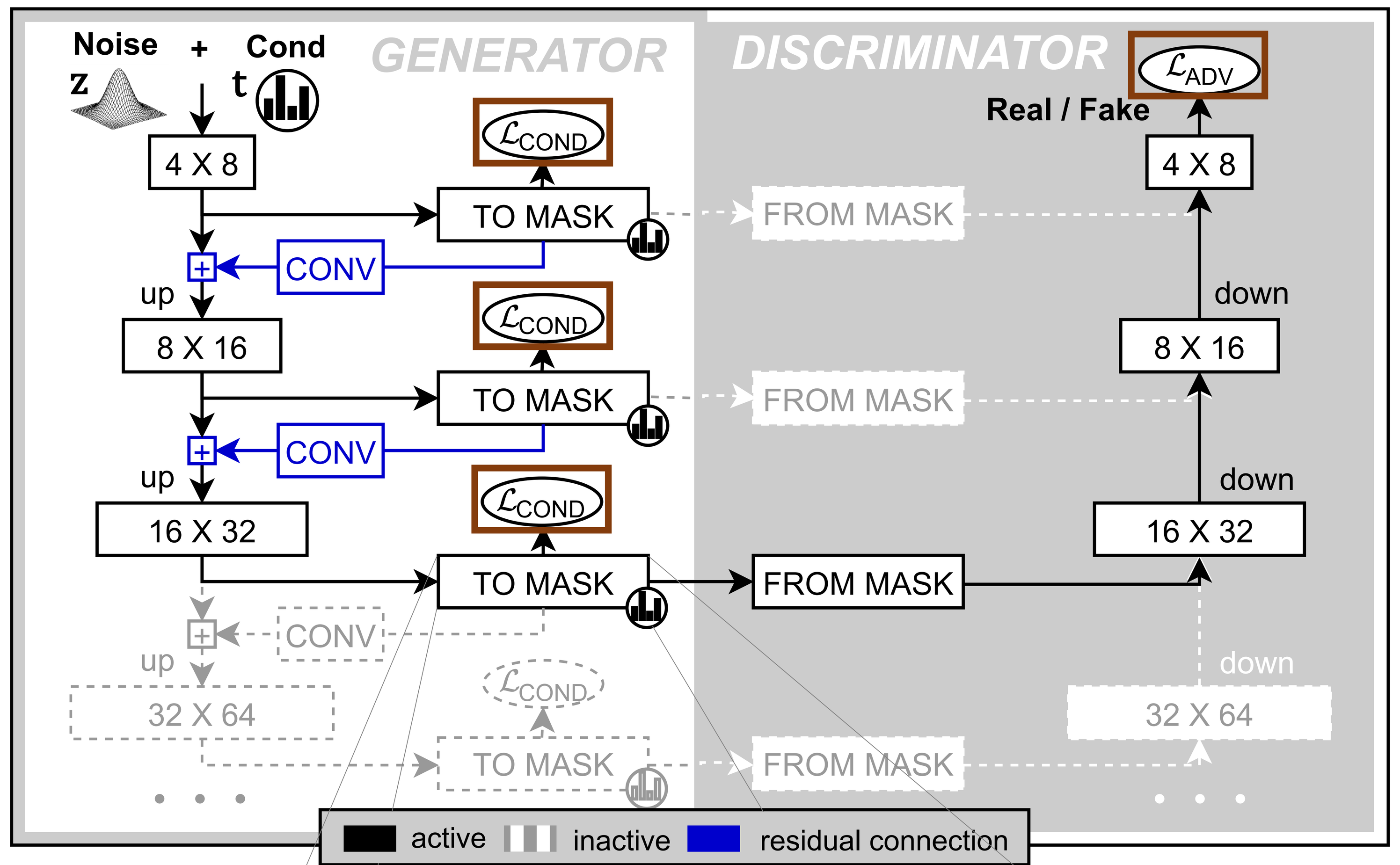
Method

Overall framework

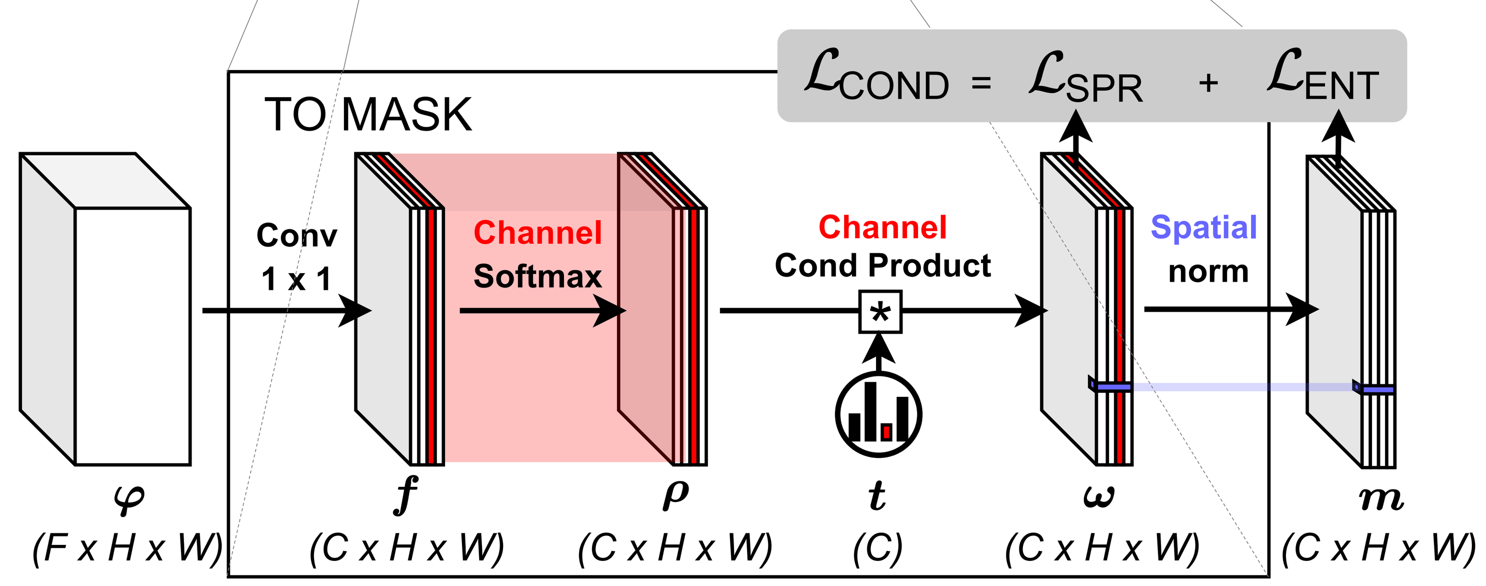
a) Learn ***palettes*** (i.e., semantic class histograms) distribution.
b) Train layout/image synthesizers individually, then end-to-end.
c) Use synthetic pairs for downstream tasks (e.g., segmentation)



Conditional layout synthesis



Semantically-assisted activation



Learning objectives

$\mathcal{L}_{ADV}$ — ① ***Create realistic layouts***

Improved WGAN loss

$\mathcal{L}_{COND}$ — ② ***Make layouts comply with target proportions***

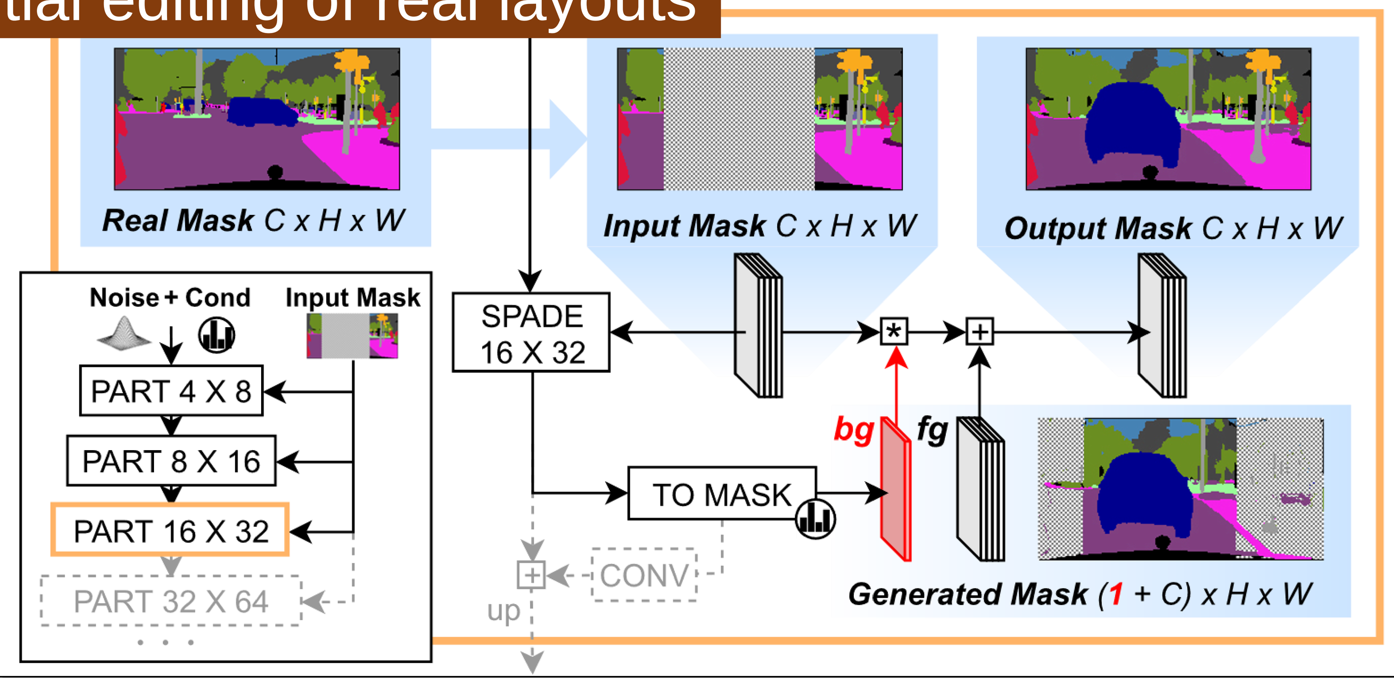
Spread loss: favor an even spatial semantic coverage

$$\mathcal{L}_{SPR} = \frac{1}{HW}\mathbb{E}_{(\boldsymbol{z},\boldsymbol{t})}\Big[\sum_{(i,j)\in\Omega} s_{i,j}\Big], \quad s_{i,j} = \big(1 - HW\textstyle\sum_{c\in[\![1,C]\!]}\boldsymbol{\omega}_{c,i,j}\big)^2$$

Entropy loss: favor a peaky class distribution at each pixel

$$\mathcal{L}_{ENT} = \frac{1}{HW}\mathbb{E}_{(\boldsymbol{z},\boldsymbol{t})}\Big[\sum_{(i,j)\in\Omega} e_{i,j}\Big], \quad e_{i,j} = -\textstyle\sum_{c\in[\![1,C]\!]}\boldsymbol{m}_{c,i,j}\ln(\boldsymbol{m}_{c,i,j})$$

Partial editing of real layouts



Results

Metrics

KL : KL-divergence between syn. and target palette
FID: Fréchet inception distance in image space
FSD: Fréchet inception distance in palette space
GAN-train: train on synthetic, test on real
GAN-test: train on real, test on synthetic

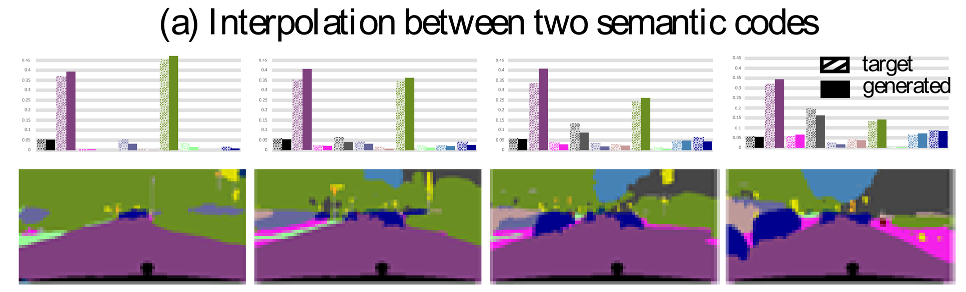
Palette *vs.* Conditional GANs

Cityscapes

Method	Layout	Image	GAN-test		GAN-train	
	KL ↓	FID ↓	mIoU*	mIoU	mIoU*	mIoU
Baseline 1	1.17	69.2	33.7	42.8	29.6	38.5
Baseline 2	0.32	69.0	35.3	46.9	30.2	39.4
Sem. Palette	**0.07**	60.7	34.6	45.7	30.6	40.1
Sem. Palette$_{e2e}$	0.08	**51.0**	**36.8**	**48.6**	**33.3**	**44.5**
Oracle	-	28.2	-	-	36.9	48.1

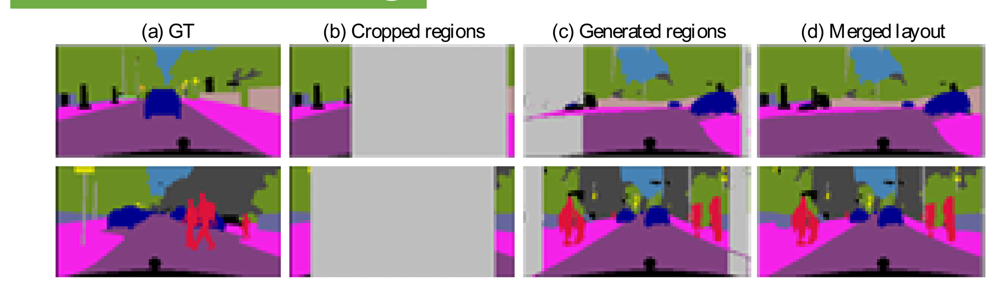
Layout-and-scene generation

(a) Interpolation between two semantic codes



(b) Diverse samples from one semantic code

Partial editing



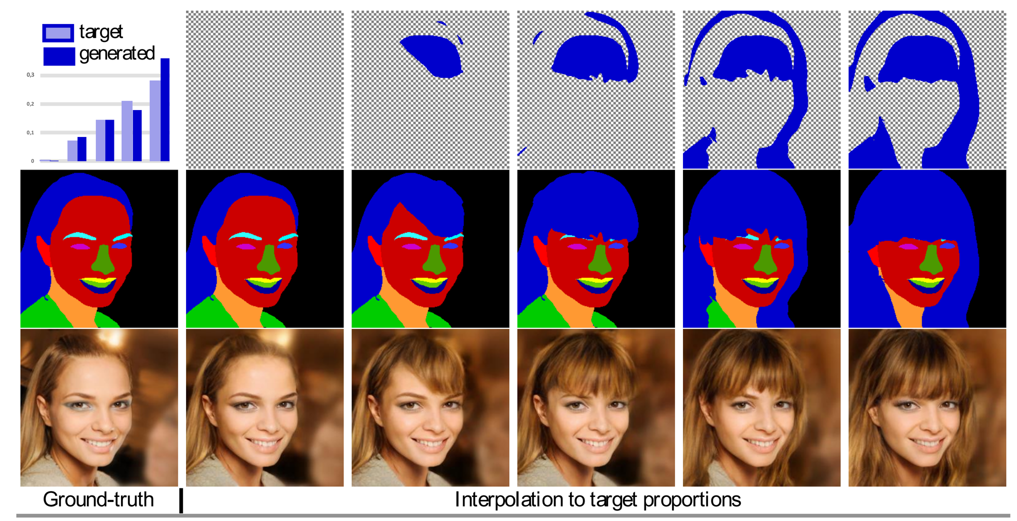
Palette *vs.* Unconditional GANs

Cityscapes

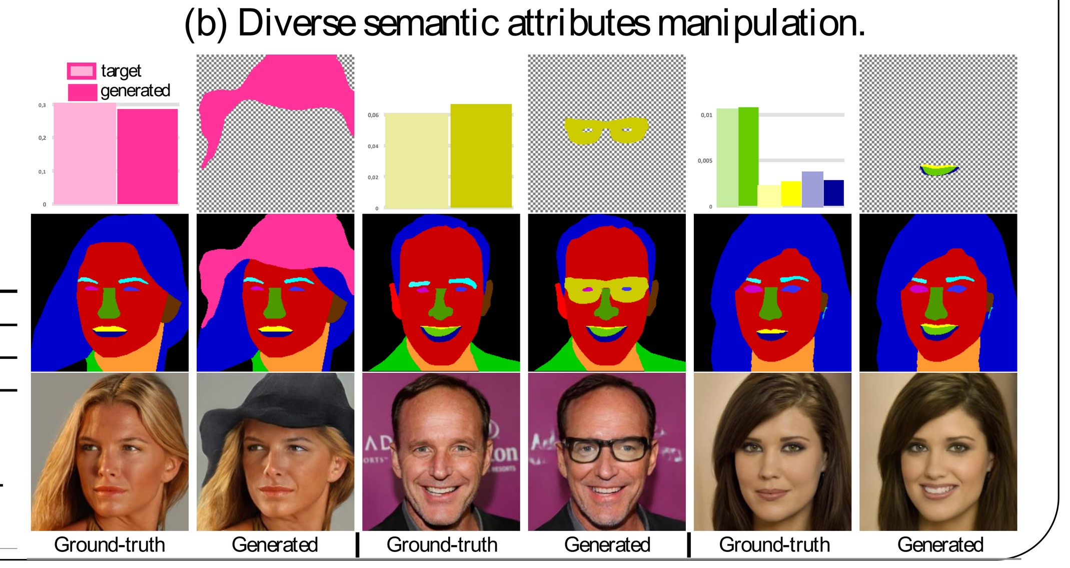
Method	Layout	Image	GAN-test		GAN-train	
	FSD ↓	FID ↓	mIoU*	mIoU	mIoU*	mIoU
PCGAN	63.8	85.7	30.4	39.0	28.2	35.7
SB-GAN	63.8	71.0	31.8	41.2	28.8	37.2
Sem. Palette	**25.3**	**60.7**	**34.6**	**45.7**	**30.6**	**40.1**
SB-GAN$_{e2e}$	20.4	61.8	34.5	44.7	29.6	37.0
Sem. Palette$_{e2e}$	**11.8**	**51.0**	**36.8**	**48.6**	**33.3**	**44.5**
Oracle	-	28.2	-	-	36.9	48.1

Face editing

(a) Hair manipulation.



(b) Diverse semantic attributes manipulation.



Data augmentation

Data	Method	(a) Cityscapes mIoU*	(a) Cityscapes mIoU	(b) IDD mIoU*	(b) IDD mIoU
Real	Baseline	36.9	48.1	33.8	43.8
Real + Semi-Syn	GauGAN	37.2 ↑0.3	48.2 ↑0.1	33.6 ↓0.2	43.5 ↓0.3
Real + Syn	SB-GAN	34.6 ↓2.3	45.5 ↓2.6	33.5 ↓0.3	43.4 ↓0.4
	Sem. Palette	38.0 ↑1.1	49.4 ↑1.3	33.8 –	43.8 –
	Sem. Palette (DA)	38.6 ↑1.7	51.6 ↑3.5	34.5 ↑0.7	44.7 ↑0.9
	Sem. Palette (Part.)	**40.7** ↑3.8	51.9 ↑3.8	**35.6** ↑1.8	**46.1** ↑2.3
	Sem. Palette (Part. + DA)	**40.7** ↑3.8	**52.6** ↑4.5	35.3 ↑1.5	45.8 ↑2.0

Conclusion

We have proposed **Semantic Palette**, a new framework for scene generation, and editing, guided by semantic proportions.

- Our novel architecture effectively accommodates class proportions while proposing plausible layouts, which then translate into realistic images.
- Semantic Palette better captures the distribution of real layouts / images than unconditional layout-and-scene GANs.
- Semantic Palette better follows target proportions and produces higher quality layouts than conditional baselines.
- Partial editing is an efficient data-augmentation strategy and opens up interesting applications like face editing.