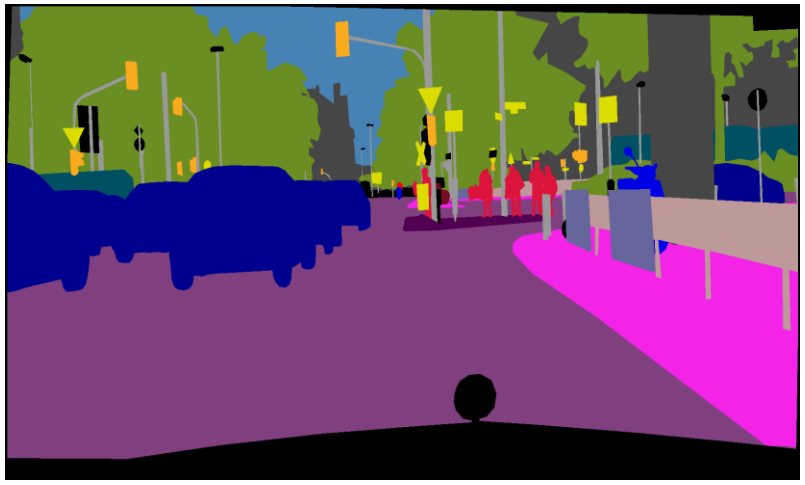
High-Resolution Image Synthesis and Semantic Manipulation with Conditional GANs

T.-C. Wang, M.-Y. Liu, J.-Y. Zhu, A. Tao, J. Kautz, B. Catanzaro
NVIDIA Corporation & UC Berkeley
In IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2018

Chen, Ting-Yin

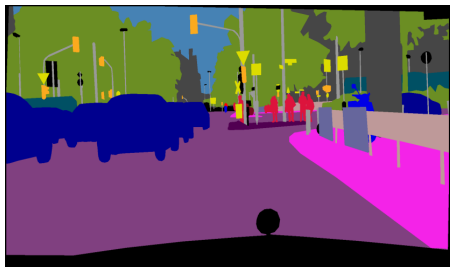
Introduction



Input
semantic label maps
2048 X 1024

Output
Synthesized image
2048 X 1024

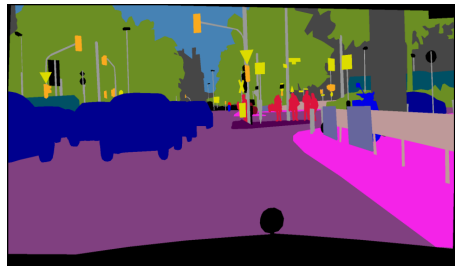
Condition GAN



semantic label maps

Generator

Image
G(s)



semantic label maps

As close as possible

Natural photo

Discriminator

scalar

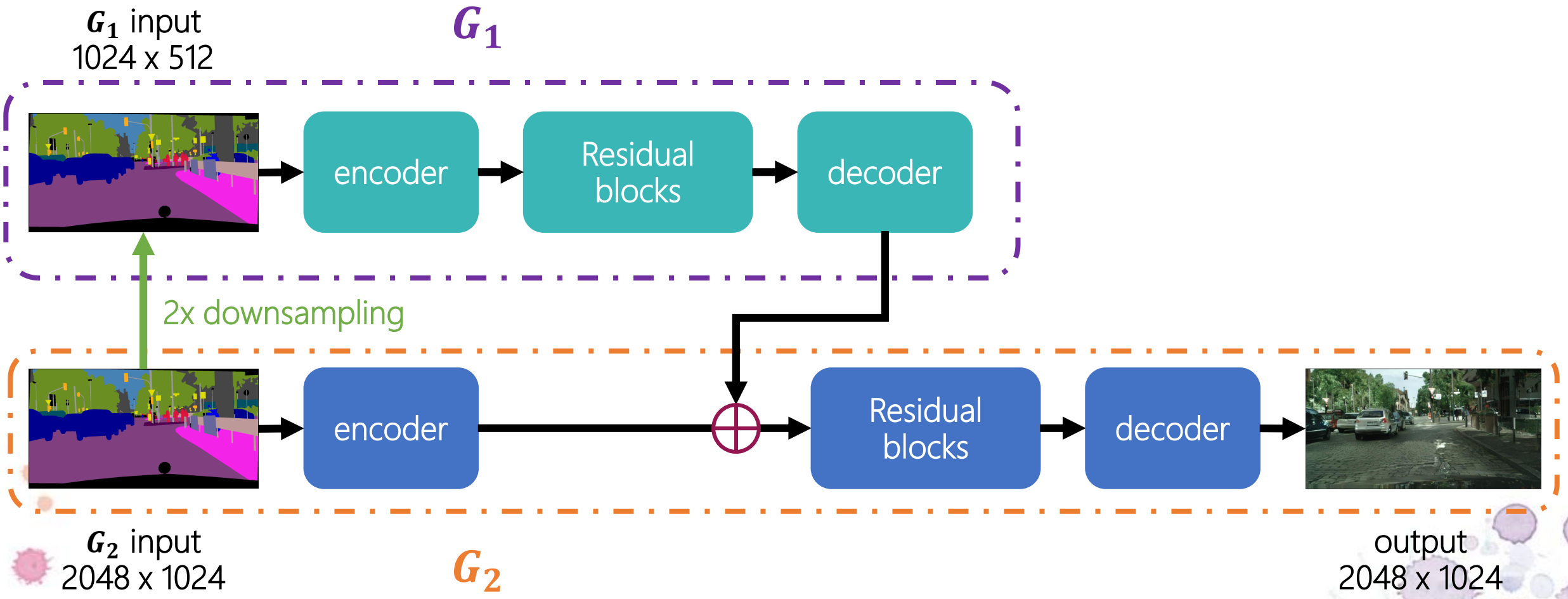
Objective function:

$$\min_G \max_D \mathcal{L}_{GAN}(G, D) = E_{(s,x)}[\log D(s, x)] + E_s[log(1 - D(s, G(s)))]$$

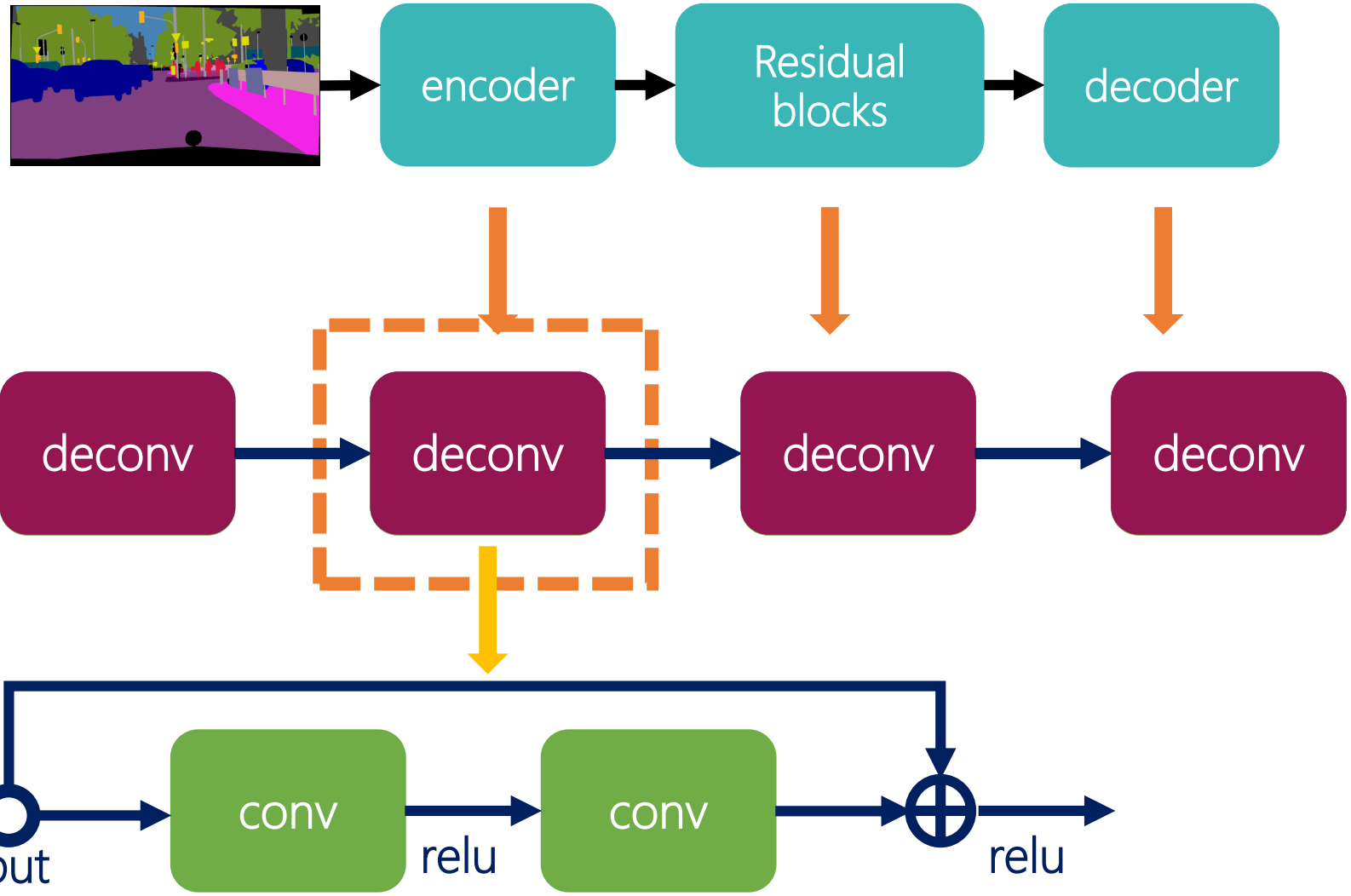
s: semantic label maps

x: natural photo

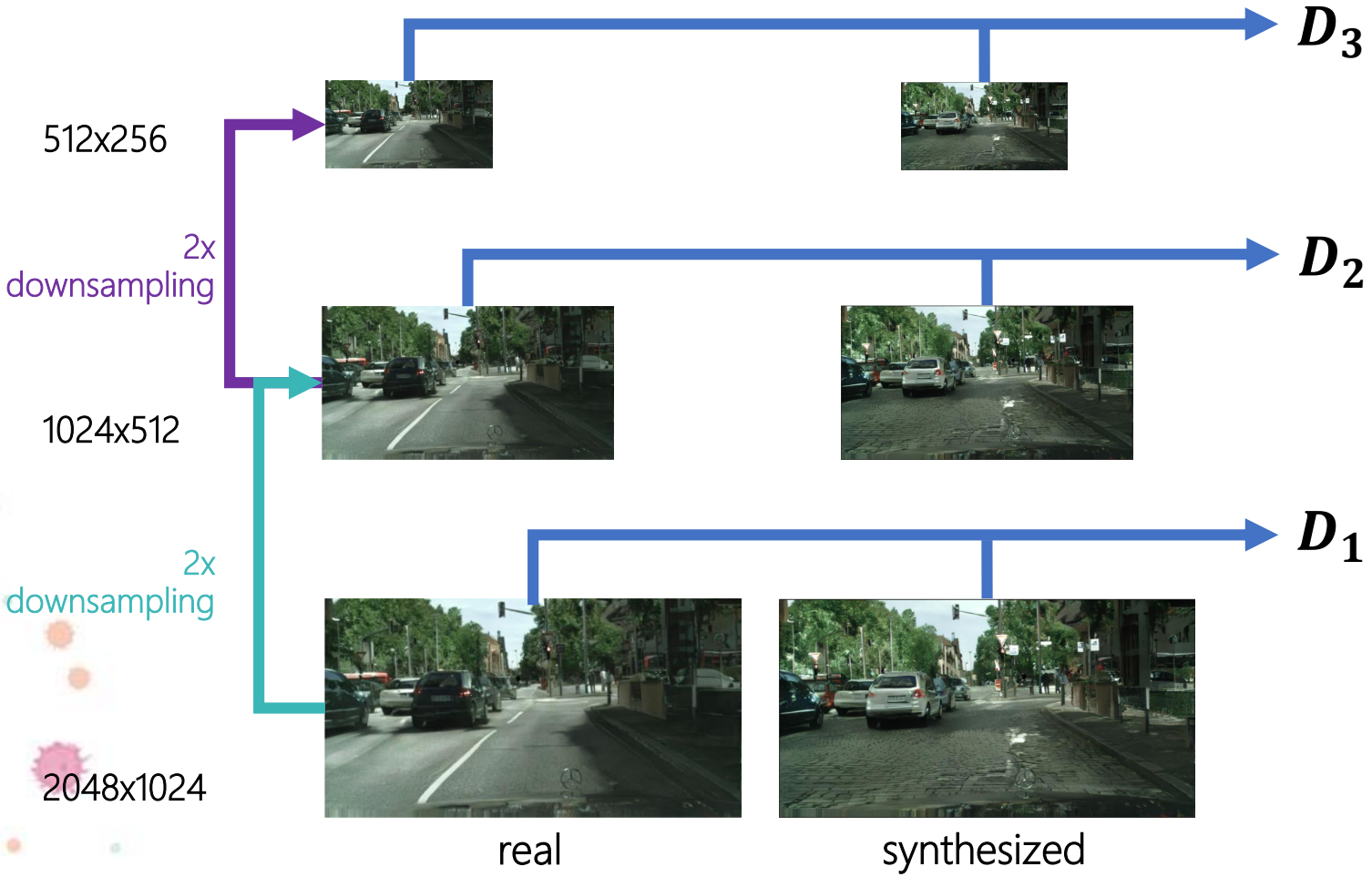
Coarse-to-fine Generator



Generator detail



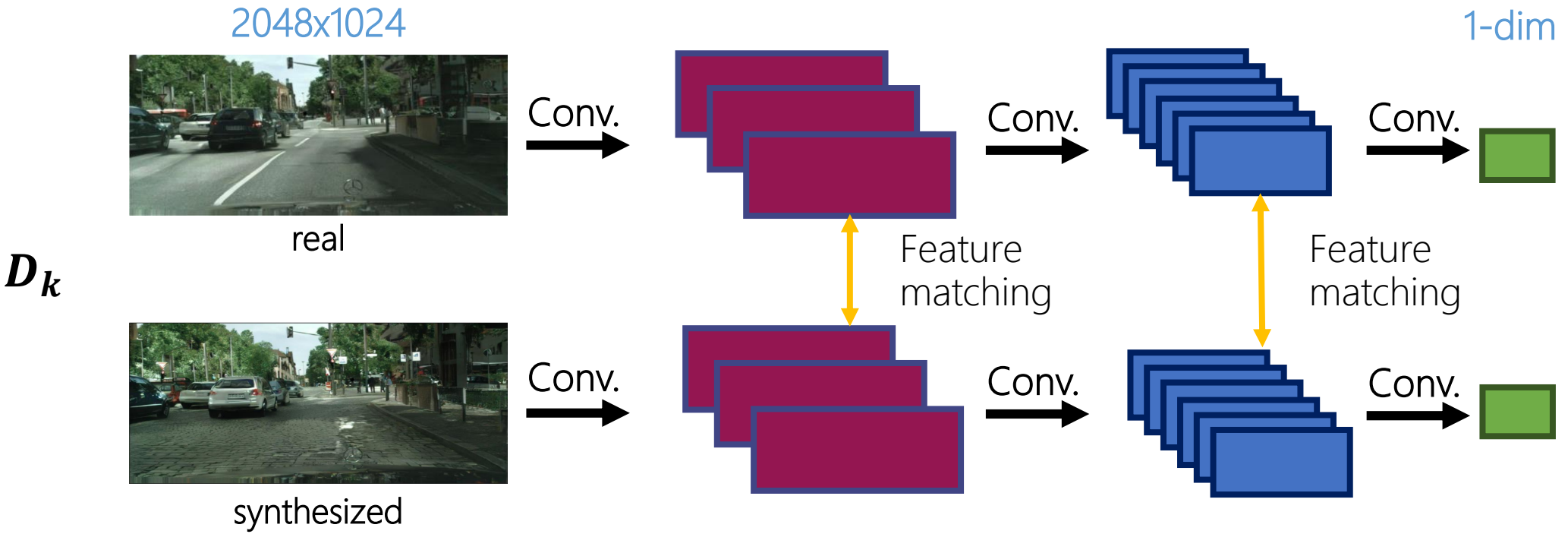
Multi-scale Discriminators



$$\mathcal{L}_{GAN}(G, D) = E_{(s,x)}[\log D(s, x)] + E_s[\log(1 - D(s, G(s)))]$$

$$\min_G \max_{D_1, D_2, D_3} \sum_{k=1,2,3} \mathcal{L}_{GAN}(G, D_k)$$

Improved adversarial loss



$$\mathcal{L}_{FM}(G, D_k) = E_{(s,x)} \sum_{i=1}^{T} \frac{1}{N_i} \left[\left\| D_k^{(i)}(s, x) - D_k^{(i)}(s, G(s)) \right\|_1 \right]$$

T: *total number of layer* $\quad$ N_i: *number of elements in* $i-th$ *layer* $\quad$ $D_k^{(i)}$: *ith layer feature extracter of* D_k

Full objective

Coarse-to-fine Generator

Multi-scale Discriminators

Improved adversarial loss

$$\mathcal{L}_{GAN}(G, D) = E_{(s,x)}[\log D(s, x)] + E_s[\log(1 - D(s, G(s)))]$$

$$\mathcal{L}_{FM}(G, D_k) = E_{(s,x)} \sum_{i=1}^{T} \frac{1}{N_i} \left[\left\| D_k^{(i)}(s, x) - D_k^{(i)}(s, G(s)) \right\|_1 \right]$$

$$\min_{G} \left(\left(\max_{D_1, D_2, D_3} \sum_{k=1,2,3} \mathcal{L}_{GAN}(G, D_k) \right) + \lambda \sum_{k=1,2,3} \mathcal{L}_{FM}(G, D_k) \right)$$