ViT_cond_NF

算法流程图

I. training, x_p = ViT (positive samples)

$$z_p = flow(x_p), \qquad z_p \sim N(0, I)$$

II. inference, x_p = ViT (positive samples)
x_n = ViT (negative samples)

① segmentation

(i, j) represents patch positions

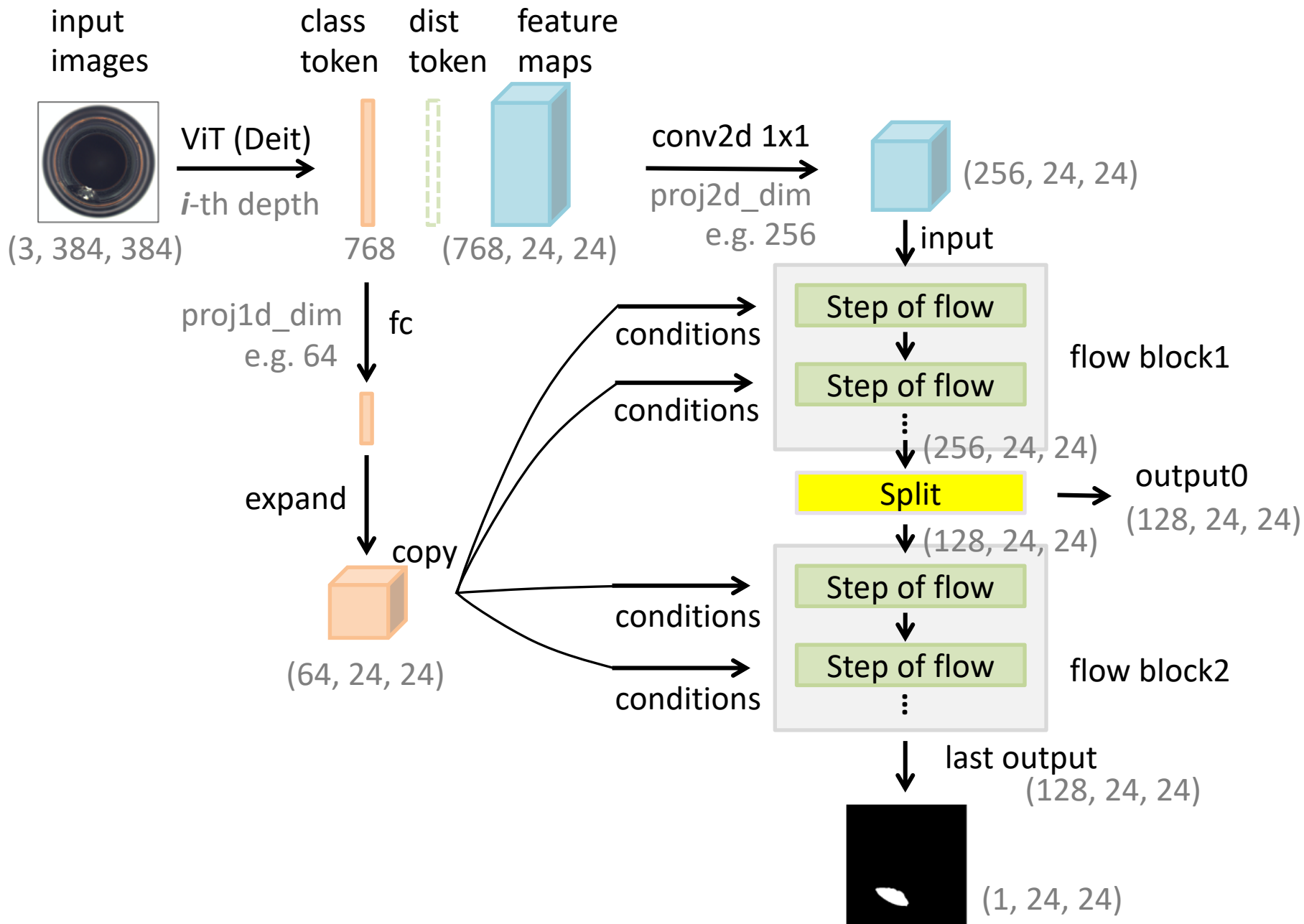$$x_p^{(i,j)} \xrightarrow{flow} z_p^{(i,j)}, \qquad ||z_p^{(i,j)}||_2 \approx 0$$

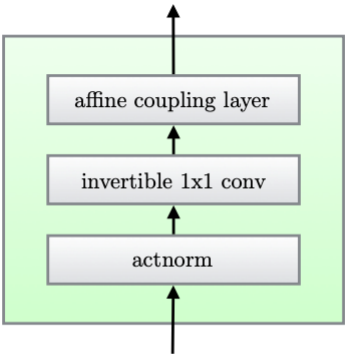$$x_n^{(i,j)} \xrightarrow{flow} z_n^{(i,j)}, \qquad ||z_n^{(i,j)}||_2 \gg 0$$

② classification

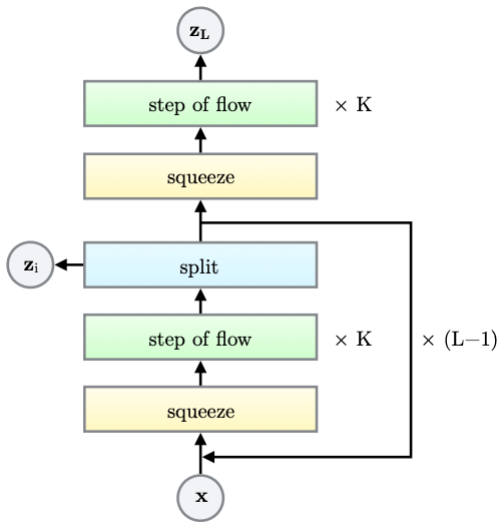*s* represents the image level anomaly score

$$s = max(Z),$$

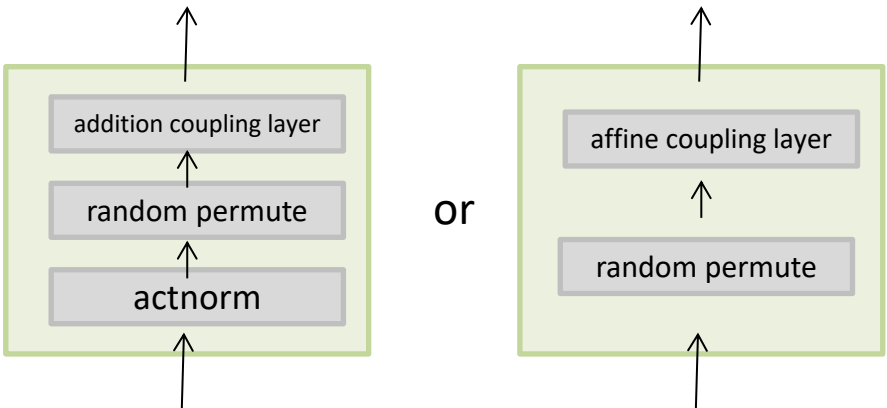Z is the set $(||z^{(1,1)}||_2, \ldots, ||z^{(i,j)}||_2, \ldots)$ of an image

input images

class token

dist token

feature maps

ViT (Deit)

i-th depth

(3, 384, 384)

768

(768, 24, 24)

conv2d 1x1

proj2d_dim

e.g. 256

(256, 24, 24)

input

proj1d_dim

e.g. 64

fc

expand

copy

(64, 24, 24)

conditions

conditions

Step of flow

Step of flow

flow block1

(256, 24, 24)

Split

output0

(128, 24, 24)

(128, 24, 24)

conditions

conditions

Step of flow

Step of flow

flow block2

last output

(128, 24, 24)

(1, 24, 24)

Glow design

(a) One step of our flow.

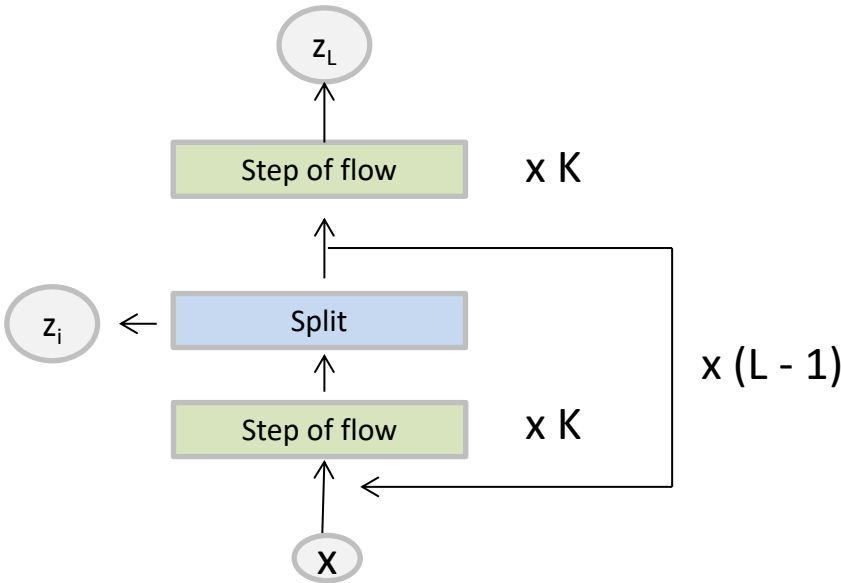(b) Multi-scale architecture (Dinh et al., 2016).

Our

or

(a) One step of our flow

x K

x (L - 1)

x K

(b) Multi-scale architecture