# Sig2Sig : Signal translation networks to take the remains of the past

SangYeon Kim, Hyunwoo Lee, Jonghee Han, Joon-Ho Kim

Samsung Research, Samsung Electronics Co.,Ltd.

Samsung Research

## Introduction

As deep learning techniques are applied to various industries, big data collection before model development has become an inevitable process. However, sensors are developed faster and replaced by new sensors, and data collection is often expensive and time consuming.Also, even if you have trained your model with great performance in the old sensor domain, you will need to modify the algorithm and retrain the model to work correctly in the new sensor domain. So, whenever a new sensor is developed, model development always incurs a significant cost. To overcome these problems, we propose a Sig2Sig (Signal-to-Signal) network based on u-net and generative adversarial networks (GAN) that convert new sensor data into existing sensor data (or vice versa). Our model aims to perform accurate sensor signal conversion between the paired signal and the generated image. Belongs to the correct class by the trained classification model. The main differences of Sig2Sig with general tasks with GAN are 1) trained on paired-dataset 2) generate a robust signal image even with noisy background 3) purpose on reusing the past model for a classification task. We tested the translated data with a action recognition model trained in the previous sensor domain.

## Main Objectives

1. A novel image translation network with attentions inside and outside of translation parts.
2. Signal translation networks that focus on important features and almost independent on noise parts.
3. The Sig2Sig model allows you to reuse pretrained classification models trained in the old sensor domain with some new sensor data. This can reduce the cost of collecting data and developing models in new domains.

## Sig2Sig

Our generative model is based on u-net as in Pix2Pix, but to make robust generative model, we have several differences with a previous model. Here, we use two types of attentions while generating images through u-net networks and generating robust images with important features after u-net networks.

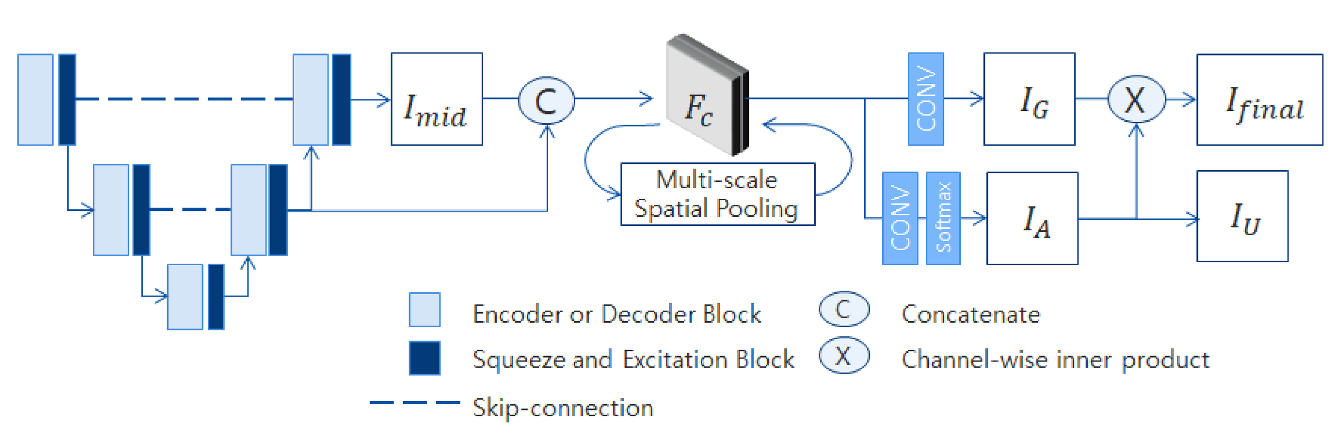



Fig. 1: Sig2Sig Model Description.

### Generative Model

· U-net with Squeeze and Excitation networks

We use attention for each layers of downsampling and upsampling among U-net networks by Squeeze and Excitation networks. Since features from downsampling layers are also used for upsampling through skip-connections, attention methods are necessary to find features that depend on inputs for better generation results than original u-net networks.

$$X^r = F(X) \tag{1}$$

$$z_k = F_{sq}(X_k^r) = \frac{1}{H \times W}\sum_{h=1}^{H}\sum_{w=1}^{W} x_k^r(h,w) \tag{2}$$

$$s = F_{ex}(z, W) = \sigma(W_2\delta(W_1 z)) \; where \; z = \{z_k\}_{k=1}^{C} \tag{3}$$

$$\tilde{x}_c = F_{scale}(u_c, s_c) = s_c \cdot u_c \tag{4}$$

$$\tilde{X} = \{\tilde{x}_k\}_{k=1}^{C} \tag{5}$$

(6)

· Multi-channel Attention and Selection

Second, we do self-attention after obtaining u-net output to generate fine images. Signal images have noise parts and sometimes they dominate the images when silent situation with only noises. Therefore, self-attention and loss weights by uncertainty maps can reduce the noise effect of the generation task.

$$F_c = concat(I_{mid}, F_{mid}) \tag{7}$$

$$F_c \leftarrow concat(F_c \cdot plup_1(F_c), ..., F_c \cdot plup_M(F_c)) \tag{8}$$

$$I_A = Softmax(F_c W_A + b_A) \tag{9}$$

$$I_{final} = \sum_{c=1}^{C} I_A^c \cdot I_G^c \tag{10}$$

(11)

· Instance Normalization

We use instance normalization instead of batch normalization, since signal data depends on background with noise and sometimes only small values can dominate the image.

$$IN(x) = \gamma\left(\frac{x - \mu(x)}{\sigma(x)} + \beta\right) \tag{12}$$

### Discriminator Model

We build each discriminator $D_A$ and $D_B$ as in Pix2Pix with spectral normalization at each layers to prevent mode collapse,

$$W_{SN}(W) := \frac{W}{\sigma(W)} \tag{13}$$

where $\sigma(W)$ is a spectral norm of weight matrix $W$. With these spectral norm matrices, we build our discriminator model similar with SNGAN.

### Losses

We use attention map $I_A$ to find out important parts of images and get uncertainty map $I_U$.

$$I_U = \sigma(W_u I_A + b_u) \tag{14}$$

where $\sigma$ is a sigmoid function for pixel-wise normalization. We apply this uncertainty map to update regular pixel-wise loss function $L_{gen}$,

$$L_{gen} = \mathbb{E}_{i=A,B}\left[\| \frac{\| x^i_{real} - x^i_{fake} \|}{I_U} + log I_U \|_1\right] \tag{15}$$

For stability training, we add reconstruction loss $L_{recon}$ for each domain,

$$L_{recon} = \mathbb{E}_{i=A,B}[\| x^i_{real} - x^i_{recon} \|_1] \tag{16}$$

where $x^i_{real}$ are input real images of domain $i = A, B$, $x^i_{recon} = G_i(x^i_{real})$.
To generate realistic image using discriminator $D_i$, we use patch loss.

$$L_{adv} = \mathbb{E}_{i=A,B}[\| D_i(x^i_{real}) \|_2 + \| 1_{patch} - D_i(x^i_{fake}) \|_2] \tag{17}$$

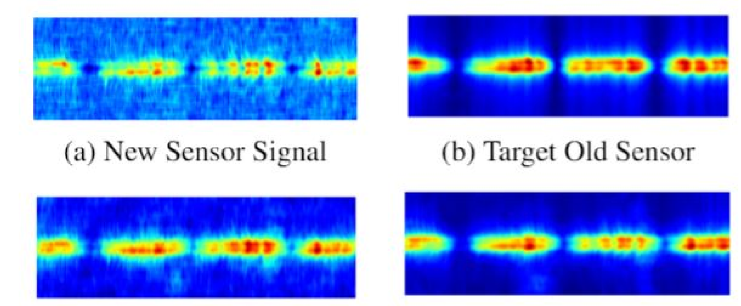
where $1_{patch}$ is a patch as same size as output of $D_i$ with ones.

Finally, we update Sig2Sig model with the total loss as following :

$$L_{total} = \lambda_{gen} L_{gen} + \lambda_{recon} L_{recon} + \lambda_{adv} L_{adv} \tag{18}$$

## Results

We used UWB(Ultra-wideband) sensor for experiment. Each signal data has 400 frames with 128 bins. For labeled data, we got 90,000 signal data with range-doppler map from signal for old domain A and 200 signal data for new domain B. To train our model, we use 900 recorded data without labels but synchronized pairs.



(a) New Sensor Signal (b) Target Old Sensor

(c) Pix2Pix (d) Sig2Sig

Fig. 2: Result images by Pix2Pix and Sig2Sig model.

The result of the Pix2Pix model looks like a blurry image of the input (new sensor signal), especially the yellow area surrounding the red area is larger than the target (old sensor signal). This can lead to incorrect classification results due to past models. On the other hand, the result of Sig2Sig is very similar to the target image than the Pix2Pix result. The Sig2Sig output has blue noise areas and finer within critical areas.

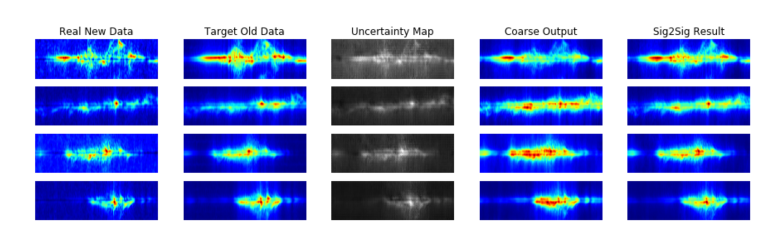


Fig. 3: Results generated by our Sig2Sig.

The uncertainty loss function on the noise area is almost constant even target image and predicted images are different and force to be similar on a focused area with less uncertainty. Therefore, after update with selection attention parts that relate to the uncertainty map, Sig2Sig results can be close to target data.

| **Old** | New | Pix2Pix | SE | IN | **Sig2Sig** |
|---|---|---|---|---|---|
| 0.92 | 0.75 | 0.85 | 0.87 | 0.88 | 0.91 |

Table 1. F1-score by the past action recognition model.

Higher f1 scores can be reached when used with the original old data. Therefore, with the Sig2Sig model, there is no degradation quality of the model and we can use the past classification model as the remains.

## Conclusions

- Only a few paired data even without labels need to train Sig2Sig.
- Sig2Sig can translate the new sensor data to old sensor data even if it is noisy.
- We can use the classification model of the past with translate new sensor data to old sensor data by Sig2Sig.