ICASSP 2023
4 - 10 JUNE, RHODES ISLAND, GREECE

Vision2Touch:Imaging Estimation of Surface Tactile Physical Properties

Jie Chen, Shizhe Zhou*

College of Computer Science and Electronic Engneering Hunan University, Changsha, Hunan, China.

湖南大學
HUNAN UNIVERSITY

INTRODUCTION

Similar to the human's multiple perception system, the robot can also benefit from cross-modal learning. The connection between visual input and tactile perception is potentially important for automated operations. However, establishing an algorithmic mapping of the visual modal to the tactile modal is a challenging task. In this work, we use the framework of GANs to propose **a cross-modal imaging method** for estimating the tactile physical properties values **based on the Gramian Summation Angular Field**, combined with visual-tactile embedding cluster fusion and feature matching methods. The approach estimates 15 tactile properties. In particular, the task attempts to predict unknown surface properties based on "learned knowledge". Our results surpass the state-of-the-art approach on most tactile dimensions of the publicly available dataset. Additionally, we conduct a robustness study to verify the effect of angle and complex environment on the network prediction performance.

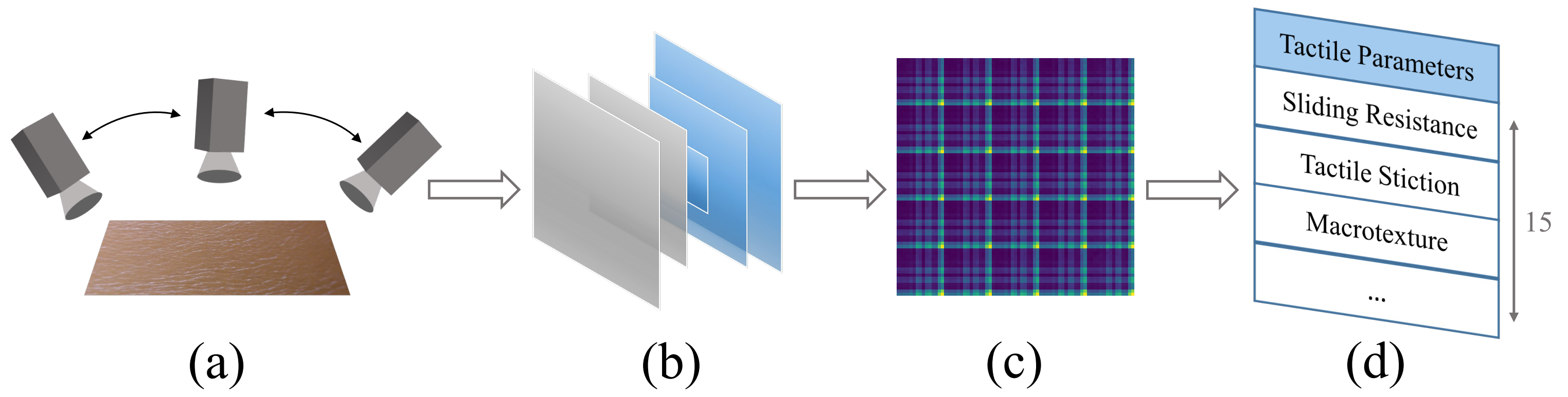

Figure 1. **Workflow.** The main concept of imaging estimation using GAN-based methods. (a) : Visual data acquisition under the standard RGB camera. (b) : Cross-modal visual-tactile generation model. (c) : Generated results with tactile information. (d) : Predicted values reduced from the generated results.

METHODOLOGY

Our idea: Modeling the transition problem from the visual to the tactile domain as an imaging estimation task based on the Gramian Summation Angular Field.

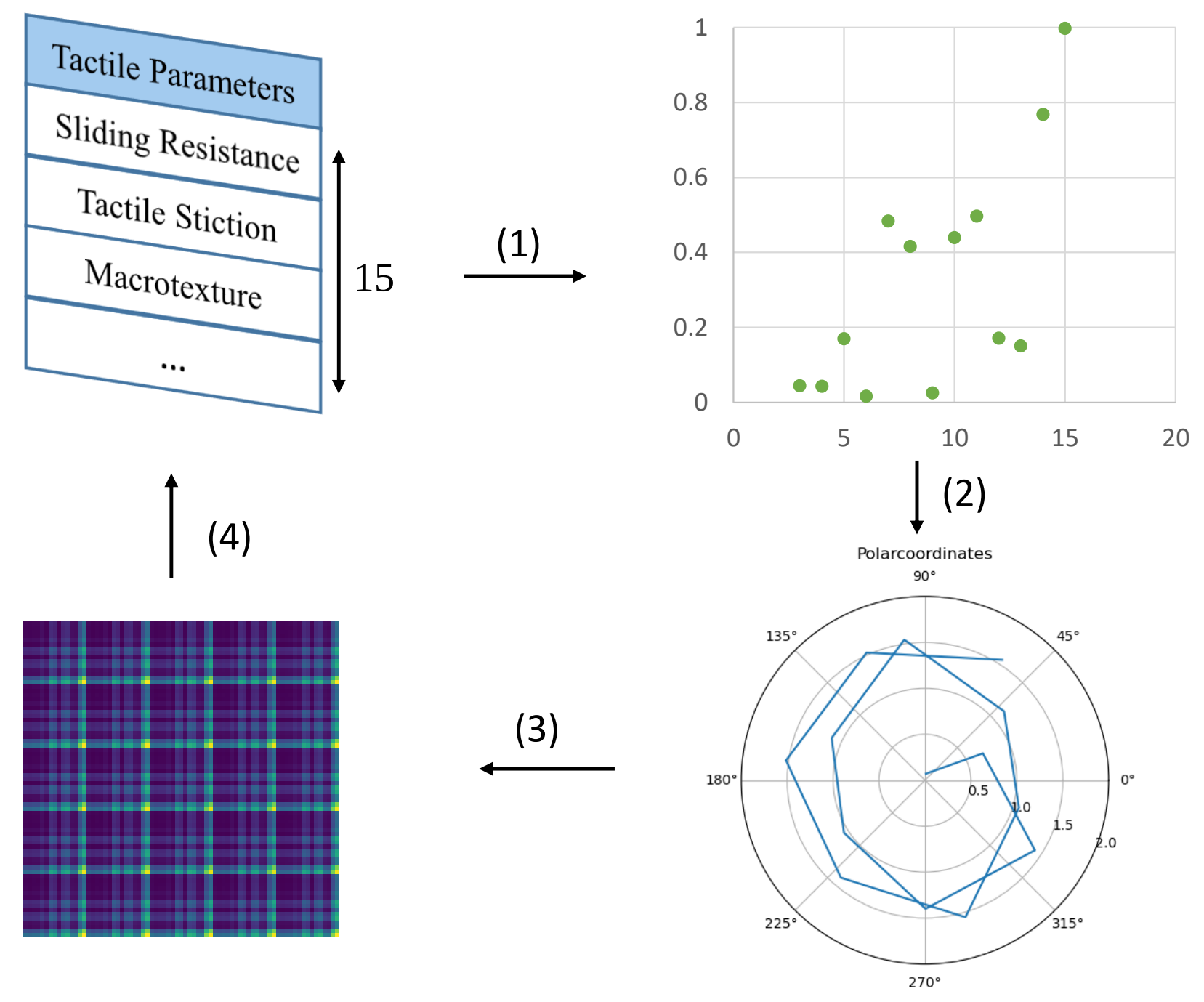

Figure 2. the Gramian Summation Angular Field workflow..

METHODOLOGY

Wang et al.[18] propose the Gramian Summation Angular Field(GASF) coding framework to encode time series into images. Motivated by these works, our work tries to **represent the tactile data as a 2D image/matrix.**

At the same time, the GASF coding **has an accurate inverse map on the [0, 1] interval**. The particular bijective property lays the foundation for GAN-based imaging estimation.

We use a polar coordinate system to represent the tactile sequence X with the equation below:

$$\phi = \arccos\left(x_i\right), 0 \leq x_i \leq 1, x_i \in X. \tag{1}$$

After converting to the polar coordinate system, we take the tactile sequence as a 1-D metric space, and by defining the inner product $\langle x, y\rangle = x \cdot y - \sqrt{1-x^2} \cdot \sqrt{1-y^2}$ we can define the GASF as follows:

$$GASF = \left[\cos\left(\phi_i + \phi_j\right)\right] \tag{2}$$

$$= X' \cdot X - \sqrt{I - \left(X'\right)^2} \cdot \sqrt{I - X^2}. \tag{3}$$

The main diagonal $GASF_{i,i}$ is the special case that contains the original value information. From the main diagonal $\{GASF_{i,i}\} = \{\cos\left(2\phi_i\right)\}$, we are allowed to precisely reconstruct the original sequences by:

$$\cos\left(\phi\right) = \sqrt{\frac{\cos\left(2\phi\right)+1}{2}}, \phi \in \left[0, \frac{\pi}{2}\right]. \tag{4}$$

NETWORK

Our model is built upon the base of the generative adversarial networks(GANs) with the GASF-based encoding framework, and trained with the additional WGAN-GP , feature-matching losses and visual-tactile embedding cluster fusion module.

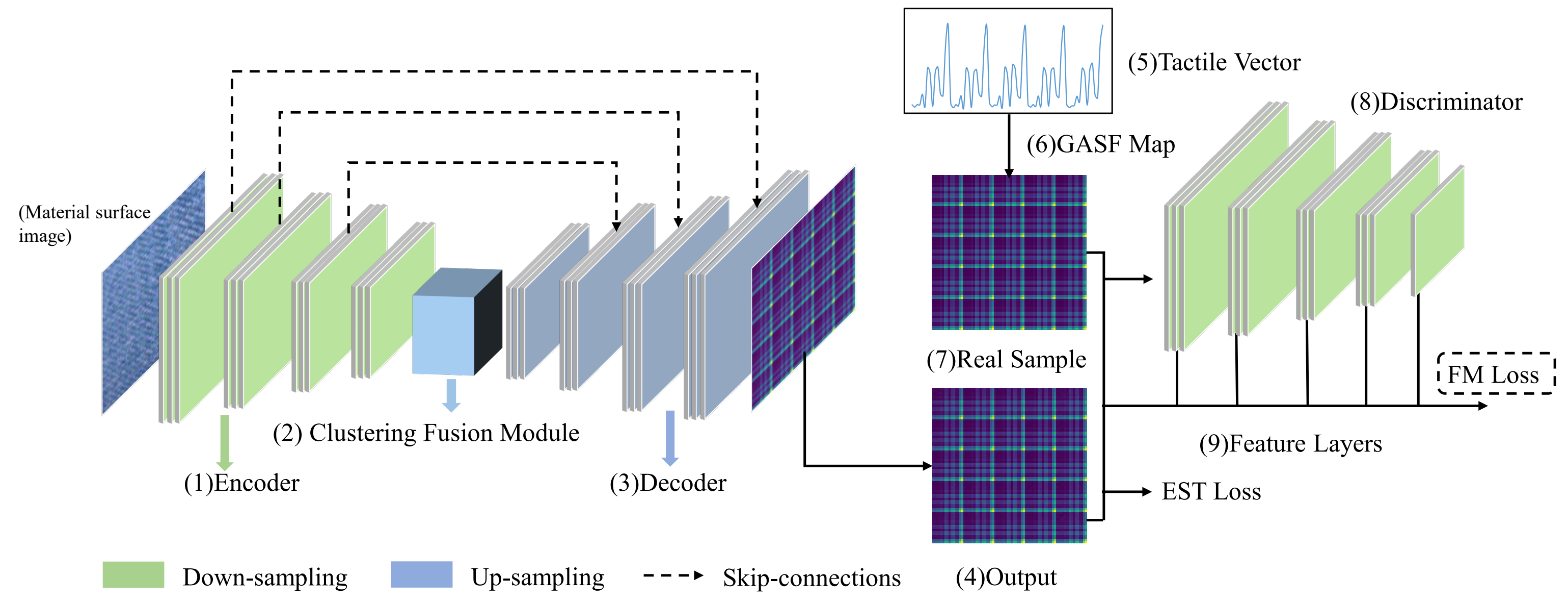

Figure 3. **The overview of the proposed framework.** Our model passes the input images to the encoder(1) and the decoder(3). The clustering Fusion module(2) is constructed with 4 residual blocks. The real image obtained from the input physical tactile vector(5) processed by the GASF coding framework(6), with the generated image(4) is input to the discriminator(8) for conditional adversarial training. We also extract the output from the feature layers of the discriminator(8) for the computation of the feature matching loss(9).

EXPERIMENTS

Dataset

We use the Surface Property Synesthesia dataset[16] as our dataset. The dataset provides a collection of surface images of 421 materials taken at different angles by a standard RGB camera under a diffuse light source. 15 surface tactile physical properties are measured with the Biotac Tactile sensing device.

Comparison Study

It can be seen from the comparison Table 1 that our proposed single-image estimation method **outperformed** the state-of-the-art results obtained in [16] on several tactile dimensions, obtaining the **best average** R^2 in the single-image estimation comparison task, and the **lowest MAE score** in all single/multiple image experiments.

Table 1. **The tactile estimation results(R^2).** The comparison experiment results are displayed. Red and blue text correspond to the first and second best scores respectively.

Properties	Baseline(single)	DEC(single)	NVS (multi)	VB-NVS (multi)	Ours
fRS	0.07	0.54	0.62	0.65	0.09
cDF	0.49	0.52	0.53	0.50	0.58
tCO	0.50	0.62	0.63	0.61	0.76
cYD	0.44	0.64	0.55	0.57	0.70
aTK	−0.46	−0.07	0.02	−0.05	0.53
mTX	0.43	0.43	0.56	0.58	0.59
cCM	0.13	0.47	0.53	0.57	0.63
cDP	0.35	0.67	0.54	0.64	0.56
cRX	0.11	0.44	0.49	0.45	0.48
mRG	0.46	0.47	0.55	0.57	0.21
mCO	0.56	0.54	0.68	0.70	0.73
uRO	0.32	0.44	0.51	0.47	0.69
tPR	0.57	0.65	0.54	0.68	0.36
uCO	0.57	0.59	0.63	0.66	0.52
fST	0.53	0.59	0.64	0.61	0.54
MeanR2	0.34	0.50	0.53	0.55	0.53
MeanMAE	6.17	5.53	5.34	5.28	4.97

CONCLUSIONS

Interactive operation of robots with environment and objects can benefit from surface property estimation to improve manipulation robustness.Such as the need for precise manipulation of different fabrics in manufacturing or ordering surfaces based on their roughness.

https://shizhezhou.github.io/

IEEE International Conference on Acoustics, Speech, and Signal Processing 2023

shizhe@hnu.edu.cn