http://www.icact.org/journal.asp

ICACT-TACT JOURNAL

Transactions on Advanced Communications Technology

Volume 8 Issue 6, November. 2019, ISSN: 2288-0003

Editor-in-Chief
Prof. Thomas Byeongnam YOON, PhD.

GIRI
Global IT Research Institute

Journal Editorial Board

Editor Guide

■ Introduction for Editor or Reviewer

All the editor group members are to be assigned as a evaluator(editor or reviewer) to submitted journal papers at the discretion of the Editor-in-Chief. It will be informed by eMail with a Member Login ID and Password.

Once logined the Website via the Member Login menu in left as a evaluator, you can find out the paper assigned to you. You can evaluate it there. All the results of the evaluation are supposed to be shown in the Author Homepage in the real time manner. You can also enter the Author Homepage assigned to you by the Paper ID and the author's eMail address shown in your Evaluation Webpage. In the Author Homepage, you can communicate each other efficiently under the peer review policy. Please don't miss it!

All the editor group members are supposed to be candidates of a part of the editorial board, depending on their contribution which comes from history of ICACT TACT as an active evaluator. Because the main contribution comes from sincere paper reviewing role.

■ Role of the Editor

The editor's primary responsibilities are to conduct the peer review process, and check the final camera-ready manuscripts for any technical, grammatical or typographical errors.

As a member of the editorial board of the publication, the editor is responsible for ensuring that the publication maintains the highest quality while adhering to the publication policies and procedures of the ICACT TACT(Transactions on the Advanced Communications Technology).

For each paper that the editor-in-chief gets assigned, the Secretariat of ICACT Journal will send the editor an eMail requesting the review process of the paper.

The editor is responsible to make a decision on an "accept", "reject", or "revision" to the Editor-in-Chief via the Evaluation Webpage that can be shown in the Author Homepage also.

■ Deadlines for Regular Review

Editor-in-Chief will assign a evalaution group(a Editor and 2 reviewers) in a week upon receiving a completed Journal paper submission. Evaluators are given 2 weeks to review the paper. Editors are given a week to submit a recommendation to the Editor-in-Chief via the evaluation Webpage, once all or enough of the reviews have come in. In revision case, authors have a maximum of a month to submit their revised manuscripts. The deadlines for the regular review process are as follows:

Evalution Procedure	Deadline
Selection of Evaluation Group	1 week
Review processing	2 weeks
Editor's recommendation	1 week
Final Decision Noticing	1 week

■ Making Decisions on Manuscript

Editor will make a decision on the disposition of the manuscript, based on remarks of the reviewers. The editor's recommendation must be well justified and explained in detail. In cases where the revision is requested, these should be clearly indicated and explained. The editor must then promptly convey this decision to the author. The author may contact the editor if instructions regarding amendments to the manuscript are unclear. All these actions could be done via the evaluation system in this Website. The guidelines of decisions for publication are as follows:

Decision	Description
Accept	An accept decision means that an editor is accepting the paper with no further modifications. The paper will not be seen again by the editor or by the reviewers.
Reject	The manuscript is not suitable for the ICACT TACT publication.
Revision	The paper is conditionally accepted with some requirements. A revision means that the paper should go back to the original reviewers for a second round of reviews. We strongly discourage editors from making a decision based on their own review of the manuscript if a revision had been previously required.

■ Role of the Reviewer

Reviewer Webpage:

Once logined the Member Login menu in left, you can find out papers assigned to you. You can also login the Author Homepage assigned to you with the paper ID and author's eMail address. In there you can communicate each other via a Communication Channel Box.

Quick Review Required:

You are given 2 weeks for the first round of review and 1 week for the second round of review. You must agree that time is so important for the rapidly changing IT technologies and applications trend. Please respect the deadline. Authors undoubtedly appreciate your quick review.

Anonymity:

Do not identify yourself or your organization within the review text.

Review:

Reviewer will perform the paper review based on the main criteria provided below. Please provide detailed public comments for each criterion, also available to the author.

- How this manuscript advances this field of research and/or contributes something new to the literature?
- Relevance of this manuscript to the readers of TACT?
- Is the manuscript technically sound?
- Is the paper clearly written and well organized?
- Are all figures and tables appropriately provided and are their resolution good quality?
- Does the introduction state the objectives of the manuscript encouraging the reader to read on?
- Are the references relevant and complete?

Supply missing references:

Please supply any information that you think will be useful to the author in revision for enhancing quality of the paperor for convincing him/her of the mistakes.

Review Comments:

If you find any already known results related to the manuscript, please give references to earlier papers which contain these or similar results. If the reasoning is incorrect or ambiguous, please indicate specifically where and why. If you would like to suggest that the paper be rewritten, give specific suggestions regarding which parts of the paper should be deleted, added or modified, and please indicate how.

Journal Procedure

Dear Author,

- **You can see all your paper information & progress.**

- **Step 1. Journal Full Paper Submission**

Using the Submit button, submit your journal paper through ICACT Website, then you will get new paper ID of your journal, and send your journal Paper ID to the Secretariat@icact.org for the review and editorial processing. Once you got your Journal paper ID, never submit again! Journal Paper/CRF Template

- **Step 2. Full Paper Review**

Using the evaluation system in the ICACT Website, the editor, reviewer and author can communicate each other for the good quality publication. It may take about 1 month.

- **Step 3. Acceptance Notification**

It officially informs acceptance, revision, or reject of submitted full paper after the full paper review process.

Status	Action
Acceptance	Go to next Step.
Revision	Re-submit Full Paper within 1 month after Revision Notification.
Reject	Drop everything.

- **Step 4. Payment Registration**

So far it's free of charge in case of the journal promotion paper from the registered ICACT conference paper! But you have to regist it, because you need your Journal Paper Registration ID for submission of the final CRF manuscripts in the next step's process. Once you get your Registration ID, send it to Secretariat@icact.org for further process.

- **Step 5. Camera Ready Form (CRF) Manuscripts Submission**

After you have received the confirmation notice from secretariat of ICACT, and then you are allowed to submit the final CRF manuscripts in PDF file form, the full paper and the Copyright Transfer Agreement. Journal Paper Template, Copyright Form Template, BioAbstract Template,

Journal Submission Guide

All the Out-Standing ICACT conference papers have been invited to this "ICACT Transactions on the Advanced Communications Technology" Journal, and also welcome all the authors whose conference paper has been accepted by the ICACT Technical Program Committee, if you could extend new contents at least 30% more than pure content of your conference paper. Journal paper must be followed to ensure full compliance with the IEEE Journal Template Form attached on this page.

➢ How to submit your Journal paper and check the progress?

Step 1. Submit	Using the Submit button, submit your journal paper through ICACT Website, then you will get new paper ID of your journal, and send your journal Paper ID to the Secretariat@icact.org for the review and editorial processing. Once you got your Journal paper ID, never submit again! Using the Update button, you can change any information of journal paper related or upload new full journal paper.
Step 2. Confirm	Secretariat is supposed to confirm all the necessary conditions of your journal paper to make it ready to review. In case of promotion from the conference paper to Journal paper, send us all the .DOC(or Latex) files of your ICACT conference paper and journal paper to evaluate the difference of the pure contents in between at least 30% more to avoid the self replication violation under scrutiny. The pure content does not include any reference list, acknowledgement, Appendix and author biography information.
Step 3. Review	Upon completing the confirmation, it gets started the review process thru the Editor & Reviewer Guideline. Whenever you visit the Author Homepage, you can check the progress status of your paper there from start to end like this, " Confirm OK! -> Gets started the review process -> ...", in the Review Status column. Please don't miss it!

Volume. 8 Issue. 6

Robust Semantic Segmentation for Street Fashion Photos

Anh H. Dang*, Wataru Kameyama†
*GITI, Waseda University, Tokyo, Japan
†Faculty of Science and Engineering, Waseda University, Tokyo, Japan
anh@aoni.waseda.jp, wataru@waseda.jp

***Abstract*—In this paper, we aim to produce the state-of-the-art semantic segmentation for street fashion photos with three contributions. Firstly, we propose a high-performance semantic segmentation network that follows the encoder-decoder structure. Secondly, we propose a guided training process using multiple auxiliary losses. And thirdly, the 2D max-pooling-based scaling operation to produce segmentation feature maps for the aforementioned guided training process. We also propose mIoU+ metric taking noise into account for better evaluation. Evaluations with the ModaNet data set show that the proposed network achieves high benchmark results with less computational cost compared to ever-proposed methods.**

***Index Terms*—semantic segmentation, street fashion photos, label pooling, mIoU+**

I. Introduction

DIFFERENT from the classic object detection and classification problem, semantic segmentation requires each pixel in the input image to be assigned to a class of objects. Fig. 1 shows examples of inputs and corresponding ground-truth labels in semantic segmentation problem for street fashion photos.

Fully convolutional neural network (FCN) for semantic segmentation [1] has laid the foundation for applying CNN into dense segmentation. Recently proposed models such as SegNet [2], DeepLabv3+ [3], and PSPNet [4] have achieved high benchmark results on data sets such as MSCOCO [5], CityScapes [6], and ADE20K [7].

ModaNet [8] is the first large-scale street fashion data set with pixel-level annotation published by S. Zheng et al.. This data set consists of 55,176 fully annotated images, where 52,377 images are for training and the remaining 2,799 images are used for validation.

In this paper, we aim to produce the state-of-the-art semantic segmentation for street fashion photos with three contributions. First, we propose a lightweight asymmetric network that follows the encoder-decoder structure. Secondly, we propose a guided training process with auxiliary training objectives. And thirdly, the 2D max-pooling-based scaling operation is proposed to produce labels to be used in one of the auxiliary training objectives. For a better evaluating segmentation result, we also propose the mIoU+ metric. Different from the conventional mIoU metric, which only counts the classification accuracy of individual pixels, segmentation noise is also taken into account in mIoU+. By both mIoU and mIoU+, the proposed network achieves the highest benchmark result in ModaNet while keeping less computational cost compared with ever-proposed methods.

The rest of the paper is organized as follows. In Section II, we introduce previous works on semantic segmentation and the use of auxiliary loss functions. In Section III, we describe our network design, auxiliary loss functions that we use to train the network, including image pyramid loss, segmentation pyramid loss, and label pooling loss. In Section IV, we describe the experimental setting, the mIoU+ metric, and the evaluation result. The paper is concluded in Section V.

Fig. 1. Samples from our custom street fashion data set. In the top row, original images are shown, and in the bottom row, corresponding segmentation ground truths are shown. The class names for each color are shown in Table I. Photos are public domain works downloaded from Pexels.com, and labels are manually annotated by the authors.

II. Related Works

A. Pre-Deep Learning Era

Semantic segmentation has been a challenge in the field of computer vision. Before the deep learning era, the state-of-the-art works have been based on Texton Forest [9] and conditional random field (CRF) [10]. CRF is still being used

Manuscript received October 19, 2019. This research is supported by funding from leaftnet Co., Ltd. in Japan.

Anh H. Dang is with GITI, Waseda University, Tokyo, Japan. (corresponding author, phone: +81-80-1367-9637, email: anh@aoni.waseda.jp)

Wataru Kameyama is with Faculty of Science and Engineering, Waseda University, Tokyo, Japan (email: wataru@waseda.jp)

as a post-process method to refine the segmentation output [1], [4], [11]–[14].

B. Fully Convolutional Neural Network

Early works on this topic mostly adopt the straight network design. The first proposed model is FCN [1]. The most important contribution of FCN is converting the fully connected classification layers of image classification networks into a 1×1 (i.e., pointwise) convolutional layers to produce pixel-level segmentation prediction. Hence, it can be implemented on top of the ever-proposed classification models such as GoogLeNet [15], VGG [16], and ResNet [17]. The authors of FCN have found that their proposal works best by using VGG-16 as the network base.

PSPNet [4] introduces the spatial pyramid pooling scheme, which results in better context-awareness in the final result. In this pyramid pooling scheme, features maps from different layers of the base network are resized and concatenated. The concatenated feature map is then used as input for a pointwise CNN to produce segmentation results.

C. Encoder-Decoder Based

Later works on the topic mostly utilize the encoder-decoder structure. Models following this approach usually yield better performance. Popular models in this category include SegNet [2] and U-Net [18]. SegNet is a CNN based autoencoder. It utilizes the indices from 2D max-pooling layers in the encoder to upscale the feature map using unpooling layers in the decoder. U-Net implements skip connections between the corresponding encoder and decoder blocks.

D. Dilated Convolutional Neural Network

In [11], F. Yu et al. propose both dilated convolutional neural network (DCNN) for semantic segmentation and a reference network design. DCNN allows the deeper layers of the network to capture the context without losing resolution. The main drawback of this design is the great demand for computational resources because the feature map is rarely down-sampled.

DeepLabv3+ [3] combines all of the above approaches and achieves state-of-the-art performance in many benchmarks.

E. Auxiliary Losses

As networks become deeper, new challenges arise. One of the most challenging problems is the vanishing gradient [19]. In this problem, the gradient becomes too small in the layers being far away from the training loss function.

At first, auxiliary losses are commonly used to overcome the problem. For example, in GoogLeNet [15], besides the main softmax classification loss at the end of the network, another two similar classification losses are added into the middle of the network. Thus, the weights of early blocks are learned mostly by gradient propagated from auxiliary losses. In the research on GANs [20], besides the usual real or fake discrimination, Chen et al. propose an auxiliary loss to discriminate the orientation of the input and output pairs to produce a more robust model. Undoubtedly, selecting the type of auxiliary objectives and their position greatly influences the performance of the network. The auxiliary training objectives also depict the type of features learned by the network. Thus, it does not guarantee that the best feature would be learned.

TABLE I
DATA SET STATISTIC

Id.	Color	Class	Inst. Count		Avg Inst. Size	
			Train	Val	Train	Val
0		Background	–	–	–	–
1		Bag	19,603	948	2.46%	2.53%
2		Belt	13,081	636	0.46%	0.44%
3		Boots	6,719	365	2.40%	2.36%
4		Footwear	37,468	1,753	0.94%	0.93%
5		Outer	22,597	1,093	7.43%	7.42%
6		Dress	13,764	662	10.46%	10.52%
7		Sunglasses	8,340	411	0.30%	0.30%
8		Pants	21,950	1,064	5.65%	5.47%
9		Top	33,131	1,544	4.79%	5.04%
10		Shorts	6,709	322	2.75%	2.83%
11		Skirts	12,953	622	6.37%	6.23%
12		Headwear	5,164	281	1.22%	1.21%
13		Scarf & Tie	4,711	284	2.85%	3.20%

Another solution to the gradient vanishing problem is using skip connections [21], [22]. In [23], skip connections are used to patch feature maps from early blocks to deeper blocks of the encoder. In [24], ResBlocks [17] are used to replace the conventional CNN blocks in both encoder and decoder, resulting in a very deep encoder-decoder based network.

Even though skip connection has become more popular due to its simplicity, it is not the replacement for auxiliary loss. Perhaps, they can be complements to each other. In [4], Zhao et al. conduct an ablation study for auxiliary loss on ResNet [17] based FCN [1]. By adding an auxiliary loss after the res4b22 residue block and weighted it appropriately, the network performance is gained by 0.94% on pixel accuracy. In [25], multiple spatially scaled versions of training labels are used as auxiliary training objectives.

III. PROPOSAL

A. Motivation

1) Problems: Two common problems of semantic segmentation are category confusion and inconspicuous segmentation [4]. Despite efforts to tackle the problems in previous researches such as [3] and [4], the problems still occur on street fashion photos, as shown in Fig. 5 in Appendix A.

In category confusion, the models fail to identify the correct class of the whole segment. For example, PSPNet fails to identify the outerwear in the image (n). And with the image (a) and (b), all the models recognize boots as an ordinary footwear. Another example of this problem is the segmentation of the image (k) by DeepLabv3+. We observe that this problem usually happens with networks that have high context-awareness.

When the above-mentioned problem is limited to local areas, it creates inconspicuous segmentation. For example, with

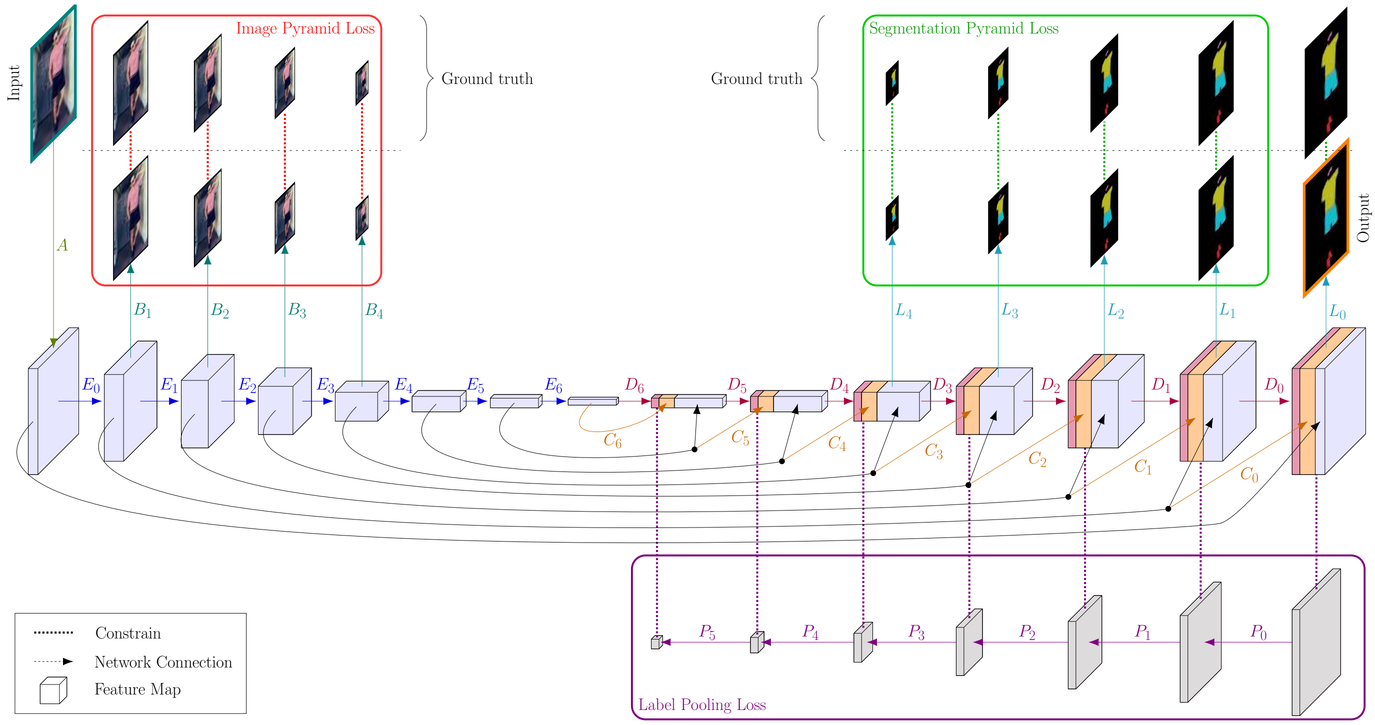

Fig. 2. Overview of the Proposed Network Structure with all three auxiliary losses. The three auxiliary losses are explained in detail in Section III-D. Ground truth for Image Pyramid Loss and Segmentation Pyramid Loss are scaled versions of the image and the ground truth segmentation. However, in Label Pooling Loss, the initial ground truth on the bottom right of the figure is one-hot vector version of the ground truth segmentation. This one-hot version of segmentation is then progressively scaled-down by $P_{[0..5]}$. This label pool feature is explained in detail in Section III-C. Moreover, in Label Pooling Lost, constrains ( ) are made only between label pool feature maps ( ) and output of decoders ( ). Network connections that are not necessary to generate output from input are ignored during inference. The detailed configuration of the whole network is described in Table II.

input image (i), SegNet detects parts of the dress as top-wear and outer-wear. With image (f), PSPNet frequently confuses between pants and skirts. Thus, it results in segmentation with a considerable amount of noise.

PSPNet [4] deliberately addresses this problem by proposing the spatial pyramid pooling module (PPM). This module is expected to increase the size of the receptive field of the network. PPM is also adopted in [3]. However, it appears that the receptive field is still limited for the street fashion problem. A possible reason is that the PPM operates on the feature map produced by a CNN head. Thus, information is already lost during the process, and the important information may not be produced simply by pooling the feature map.

2) Direction: We observe that, for street fashion photo, the above-mentioned problems can be eliminated by knowing whatever a type of apparels is presented in the whole image. For example, in the image (d) of Fig. 5, there are only two types of apparels presented in the image that are dress and pants (a small dark gray area under the model's left arm as in the ground truth image). Thus, if a network only considers dress and pants for the segmentation result, the problem of class confusion and inconspicuous would greatly be reduced.

On the one hand, it is uncomplicated to create a separated model to detect whatever the type of clothes is presented in an image. On the other hand, it is not efficient to create and train separated networks to solve a single problem. Therefore, we merge two types of networks into one and further extend the concept of apparel detector to all of the scales.

3) Implementation: Based on the encoder-decoder structure, we first set the length of the network so that the feature map at the end of the decoder is 1×1. It is to ensure the high context awareness of the network. Secondly, at every scale of the decoder, we expect the network to produce a prediction to indicate whatever the type of clothes is presented in the receptive field of the corresponding pixel of the feature map, i.e., the network first detects the presence of apparel type over the whole image, and then refines it until reaching the required resolution. Ground truth for such prediction can be produced by applying 2D max-pooling on a one-hot vector form of the original ground truth. This process is explained in detail in Section III-C.

B. Network Structure

Fig. 2 shows the structure of our proposed network. It comprises two main parts: encoder and decoder. Both the encoder and decoder parts consist of 7 CNN blocks (E_i and D_i blocks, where $i \in [0..6]$). Feature map is downscaled every time it is processed by an encoder block, and correspondingly upscaled every time it is processed by a decoder block. We organize this network into 7 different levels based on 7 different scales of the feature maps.

Similar to U-Net, skip connections with identity function are implemented between encoder and decoder blocks of the same level (black arrows as in Fig. 2). However, in our proposal, the feature map produced by an encoder also leaks

TABLE II
NETWORK PARAMETERS

Block	Output	Filters	Kernel	St.[a]	Pd.[b]	Ac./Op.[c]
A	224×224	32	5×5	1	2	ReLu
		64	3×3	1	1	ReLu
		32	1×1	1	0	ReLu
B_1	112×112	64	3×3	1	1	ReLu
		3	3×3	1	1	Sigmoid
B_2	56×56	– ” –	– ” –	– ” –	– ” –	– ” –
B_3	28×28	– ” –	– ” –	– ” –	– ” –	– ” –
B_4	14×14	– ” –	– ” –	– ” –	– ” –	– ” –
E_0	112×112	64	4×4	2	1	ReLu
		128	3×3	1	1	ReLu
		64	1×1	1	0	ReLu
E_1	56×56	128	– ” –	– ” –	– ” –	– ” –
E_2	28×28	256	– ” –	– ” –	– ” –	– ” –
E_3	14×14	512	– ” –	– ” –	– ” –	– ” –
E_4	7×7	1024	– ” –	– ” –	– ” –	– ” –
E_5	3×3	1024	3×3	3	1	ReLu
		2048	3×3	1	1	ReLu
		1024	1×1	1	0	ReLu
E_6	1×1	1024	3×3	1	0	ReLu
		2048	1×1	1	0	ReLu
		1024	1×1	1	0	ReLu
C_0	112×112	128	3×3	1	1	ReLu
	224×224	1	2×2	2	0	UnPool
C_1	112×112	– ” –	– ” –	– ” –	– ” –	– ” –
C_2	56×56	– ” –	– ” –	– ” –	– ” –	– ” –
C_3	28×28	– ” –	– ” –	– ” –	– ” –	– ” –
C_4	14×14	– ” –	– ” –	– ” –	– ” –	– ” –
C_5	3×3	128	3×3	1	1	ReLu
	7×7	1	3×3	3	1	UnPool
C_6	1×1	128	1×1	1	0	ReLu
	3×3	1	3×3	3	0	UnPool
D_0	112×112	128	3×3	1	1	ReLu
		14	3×3	1	1	Sigmoid
	224×224	1	2×2	2	0	UnPool
D_1	112×112	– ” –	– ” –	– ” –	– ” –	– ” –
D_2	56×56	– ” –	– ” –	– ” –	– ” –	– ” –
D_3	28×28	– ” –	– ” –	– ” –	– ” –	– ” –
D_4	14×14	– ” –	– ” –	– ” –	– ” –	– ” –
D_5	3×3	128	3×3	1	1	ReLu
		14	3×3	1	1	Sigmoid
	7×7	1	3×3	3	1	UnPool
D_6	1×1	128	1×1	1	0	ReLu
		14	1×1	1	0	Sigmoid
	3×3	1	3×3	1	0	UnPool
P_0	56×56	1	2×2	2	0	MaxPool
P_1	28×28	– ” –	– ” –	– ” –	– ” –	– ” –
P_2	14×14	– ” –	– ” –	– ” –	– ” –	– ” –
P_3	7×7	– ” –	– ” –	– ” –	– ” –	– ” –
P_4	3×3	– ” –	3×3	3	1	– ” –
P_5	1×1	– ” –	3×3	1	1	– ” –
L_0	224×224	128	3×3	1	1	ReLu
		14	3×3	1	1	Softmax
L_1	112×112	– ” –	– ” –	– ” –	– ” –	– ” –
L_2	56×56	– ” –	– ” –	– ” –	– ” –	– ” –
L_3	28×28	– ” –	– ” –	– ” –	– ” –	– ” –
L_4	14×14	– ” –	– ” –	– ” –	– ” –	– ” –

[a]Stride, [b]Padding, [c]Activation/Operation

into the next level of the decoder. To adapt the feature map into the larger scale, we employ CNN - 2D unpooling blocks C_i. In our network, encoder blocks scale down the feature map by utilizing CNN with stride 2 instead of using 2D max-pooling operation. Thus, different from SegNet [2], the 2D unpooling layer in our network doesn't utilize the pooling indices.

As mentioned, in this network, we expect decoder blocks to produce the prediction on the presence of a class within the whole image and then gradually refine the prediction result. Therefore, all the decoder blocks have the same design that output only 14 channels feature map, which is the number of segmentation classes of ModaNet (13 classes plus background, as shown in Table II).

Element-wise sigmoid function is used as the activation function for D_i blocks as follows.

$$d_i = \frac{1}{1 + \exp(-d_i')} \tag{1}$$

Where d_i is the output of D_i block, and d_i' is the pre-activation value of D_i.

Segmentation prediction is produced by L_0 block. In this network, besides L_0, there are another 4 L_i blocks where $i \in [1..4]$. These blocks produce smaller-scale versions of the segmentation prediction. In general, the scale of the prediction produced by L_i block is 2^{-i}. The input of L_i block is the concatenation of feature maps output from C_i, D_i and E_i blocks. Pixel-wise softmax is used to produce the output of L_i blocks as follows.

$$l_{ij} = \frac{\exp\left(l_{ij}'\right)}{\sum_{k=1}^{K} \exp\left(l_{ik}'\right)} \tag{2}$$

Where l_{ij} denotes the value of channel j of the feature map produced by L_i, l_{ij}' denotes the pre-activation value of l_{ij}, and K denotes the total number of channels which also is the number of segmentation classes.

From level 1 to level 4, different scales of the input image are reconstructed by B_i blocks where i is the level number. The input of B_i block is the feature map e_{i-1} produced by E_{i-1} block. All the B_i blocks reconstruct the input at the scale of 2^{-i}. Thus, all the outputs from B_i blocks create a spatial scale pyramid of the input image. Element-wise sigmoid, as in (1) is used as the activation function for B_i layers. Thus, different from works such as [25] and [26], we are not using the image pyramid as input but as auxiliary training objective.

C. Label Pooling

Previous works involving multiple-scale inputs or outputs only consider spatial scaling. In [25], they are used as an auxiliary training objective. In [26], they are used to create multi-scale fusion features. In [4] and [3], the network is trained with different spatial scaled versions of input and output to produce more robust features.

However, with spatial scaling, details from the original input eventually are lost at smaller scales. To avoid such problems, instead of spatially scaling the label, we use max-pooling operation on the one-hot label vector to produce multiple scales of labels. As such, the existence of a segment is

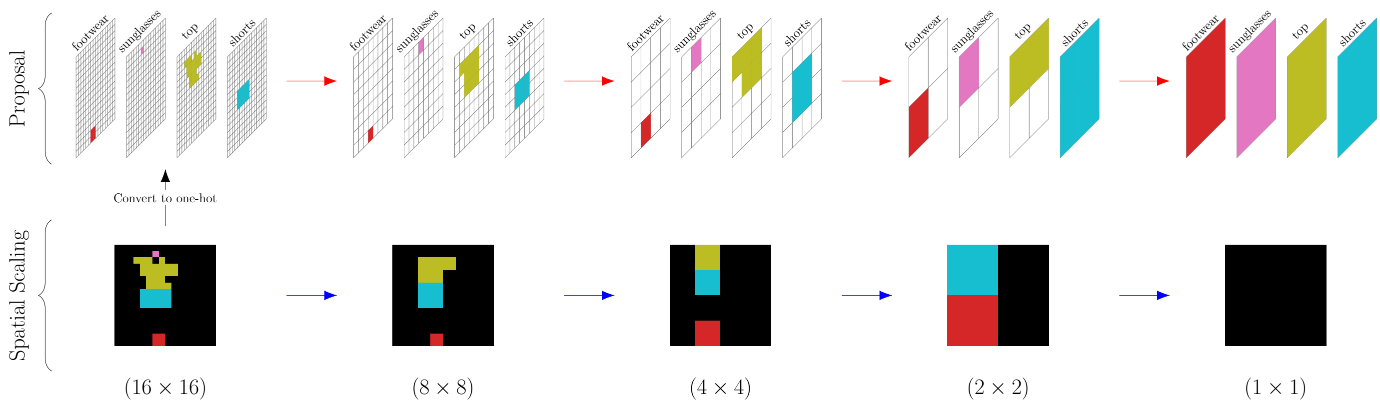

Fig. 3. Comparison between proposed 2D max-pooling-based label scaling and conventional label scaling. With conventional label scaling, the label is progressively scaled-down using nearest-neighbor interpolation (blue arrows). In our proposal, the original label (bottom left) is first converted to one-hot vectors (top left) and then progressively scaled-down by 2D max-pooling operation (red arrows). The segmentation color codes in the label are described in Table I. On the top row, classes rather than footwear, sunglasses, top, and shorts are ignored.

preserved even in the smallest scale. This process is illustrated in Fig. 3.

In Fig. 3, the spatial scaling operation makes the existence of segmentation vanished. After the first scale down operation, the segmentation of sunglass class has vanished. From the scale of 4×4 to 2×2, the top segment has vanished. By the time of scaling down to 1×1 pixel, all the segmentations vanished. On the other hand, the proposed label pool features retain all of the segmentation even at the smallest scale.

Shown in Fig. 2, D_i blocks are guide-trained by the result of P_i blocks where i is the level number, and P_i blocks are 2D max-pooling operation on the input label. Their configuration is shown in Table II. This is to avoid the detail loss when scaling down the label. Also shown in Table II, the strides of P_i are matched with the strides of D_i and E_i blocks. Furthermore, the kernel size of P_i is also matched with the kernel size of D_i.

D. Training Objectives

With the segmentation prediction l_0 coming from L_0, we utilize pixel-wise cross-entropy as a training objective.

$$H(t, l_0) = \frac{1}{N} \sum_{i=0}^{N-1} \sum_{j=0}^{K-1} t_{ij} \log(l_{0ij}) \tag{3}$$

Where t is the ground-truth, t_{ij} is the j-th channel of the i-th pixel of t, l_{0ij} is the j-th channel of the i-th pixel of l_0, K is the number of segmentation class, and N is the total number of pixel in the output. Besides this conventional training criteria, we introduce the three auxiliary training objectives as follows.

1) Image Pyramid Loss (IPL): Different from popular works, we do not utilize multi-scaled input to reinforce the training process. Instead, we expect the network to reconstruct scaled versions of the input using feature maps produced by encoder blocks. Thus, additional scaled inputs and processing are not required during the inference process. We penalize the difference between output b_i from B_i block and the input image x_i by binary cross-entropy loss as follows.

$$H(x_i, b_i) = -\frac{1}{N} \sum_{j=0}^{N-1} \Big(x_{ij} \cdot \log(b_{ij}) + (1 - x_{ij}) \cdot \log(1 - b_{ij}) \Big) \tag{4}$$

Where x_i and b_i are input image and reconstructed image at scale 2^{-i} with $i \in [1..4]$, x_{ij} and b_{ij} are the j-th element of x_i and b_i, N is the total number of elements in x_i and b_i (i.e. number of pixel $\times$ number of color channels). We use binary cross entropy (i.e. log loss) as the error function. Then, the image pyramid loss is calculated by:

$$IPL = \frac{1}{4} \sum_{i=1}^{4} H(x_i, b_i) \tag{5}$$

2) Segmentation Pyramid Loss (SPL): It is the average of the cross-entropy between segmentation and ground truth across different scales.

$$SPL = \frac{1}{4} \sum_{i=1}^{4} H(t_i, l_i) \tag{6}$$

Where $H(\cdot)$ is binary cross-entropy loss similar to (4), t_i and l_i are ground truth and predicted segmentation at scale 2^{-i} with $i \in [1..4]$.

3) Label Pooling Loss (LPL): It is the average of binary cross-entropy loss between label pool features and output of decoders across different scales as follows.

$$LPL = \frac{1}{6} \sum_{i=1}^{6} H(p_i, d_i) \tag{7}$$

Where $H(\cdot)$ is binary cross-entropy loss similar to (4), p_i and d_i are ground truth and prediction of label pool feature at scale 2^{-i}.

The final loss is calculated by averaging all the above mentioned losses as follows:

$$loss = \frac{1}{4} \left(H(t, l_0) + IPL + SPL + LPL \right) \tag{8}$$

Fig. 4. Illustration of segmentation results. Photos are public domain works downloaded from Pexels.com. Label are manually annotated by the authors.

IV. Evaluation

Using the ModaNet data set, we compare our model with U-Net [18], PSPNet [4], SegNet [2], and DeepLabv3+ [3].

A. Data Set

We split the original training set into new training and evaluation sets. The new evaluation set consists of 2,400 images, and the new training set consists of the remaining 49,977 images.

The randomized splitting process is constrained so that there are at least 280 instances of each class available in the evaluation set to ensure the quality of evaluation. The statistic of training and validation data sets are shown in Table I.

B. Data Augmentation

We train and evaluate all the networks with input and output sizes of 224×224. To make the network more robust, the following pipeline is used for data augmentation:

1) Random horizontal flipping
2) Random expanding with a max expansion ratio of 1.5. In this step, black bars of random size t, b, l and r are padded into the original image so that $l + r \leq 0.5 \times w$ and $t + b \leq 0.5 \times h$, where w and h are width and height of the input of this step, t and b are the sizes of black bars padded on the top and bottom of the image, and l and r are the sizes of black bars padded on the left and right side of the image.
3) Randomly cropping the image with the scale ratio range $(0.5, 1]$ and aspect ratio range $[\frac{3}{4}, \frac{4}{3}]$. Thus, width w and height h of the cropping window are randomized so that:
   - $w \leq w_0$ and $h \leq h_0$
   - $\frac{3}{4} \leq \frac{w}{h} \leq \frac{4}{3}$
   - $0.5 \times (w_0 \times h_0) < w \times h \leq w_0 \times h_0$

   Where w_0 and h_0 are the width and height of the input of this step. Hence, the top left corner of the cropping window (x, y) must satisfy the following conditions:
   - $0 \leq x \leq w_0 - w$
   - $0 \leq y \leq h_0 - h$
4) Adding Gaussian noise with mean $\mu = 0$ and standard deviation $\sigma = 25.5$. Thus, the output image is $x = \min(255, \max(0, x + \mathrm{G}(\mu, \sigma)))$ where $\mathrm{G}(\cdot)$ is the Gaussian function.
5) Resize to $224 \times 224 \times 3$

We then normalize the input image by scaling pixel value into $[0, 1]$ range. Since the label is an image containing pixel-level segmentation of the input image, it also needs to be augmented correspondingly, except for the step 4. Furthermore, nearest-neighbor sampling must be used in all the steps that involve interpolation to preserve class information.

C. Metrics

1) mIoU: We utilize intersection over union (IoU, i.e., Jaccard distance) as the performance metric. We first compute the IoU of individual class as follows.

$$IoU_i = \frac{1}{N} \sum_{j=0}^{N-1} \frac{|T_{ij} \cap L_{ij}|}{|T_{ij} \cup L_{ij}|} \tag{9}$$

Where IoU_i is the IoU score of class i, N is the total number of photos in the data set, T_{ij} is the set of all the

TABLE III
IoU OF INDIVIDUAL CLASSES

	A	B	C	D	E	F	G	H	Unet	DLv3+[a]	PSPNet	SegNet
Background	0.979	0.979	0.978	0.979	*0.979*	**0.980**	0.978	0.979	0.979	0.975	0.977	0.955
Bag	0.690	0.692	0.691	0.694	0.693	0.689	0.707	0.694	*0.701*	0.674	**0.708**	0.429
Belt	*0.465*	0.462	0.467	0.442	0.456	0.454	0.454	0.440	**0.479**	0.394	0.415	0.205
Boots	0.556	**0.577**	*0.570*	0.543	0.575	0.557	0.569	0.559	0.561	0.535	0.567	0.369
Footwear	0.630	0.624	0.629	0.628	**0.638**	0.638	0.631	0.622	*0.637*	0.586	0.555	0.452
Outer	0.642	*0.644*	0.623	0.629	0.639	0.651	0.627	0.638	0.623	0.625	**0.665**	0.438
Dress	0.669	*0.668*	0.633	0.663	0.651	0.657	0.651	0.649	0.594	0.660	**0.689**	0.462
Sunglasses	0.634	0.611	0.652	0.606	*0.650*	0.625	0.646	0.621	**0.675**	0.531	0.534	0.321
Pants	*0.805*	**0.810**	0.793	0.790	0.802	0.808	0.801	0.792	0.787	0.761	0.800	0.683
Top	0.647	0.650	0.625	0.641	*0.653*	0.660	0.641	0.652	0.624	0.610	**0.679**	0.478
Shorts	0.686	**0.718**	0.686	0.697	0.697	0.692	0.688	0.708	0.662	0.715	*0.711*	0.487
Skirts	0.661	*0.674*	0.659	0.671	0.646	0.673	0.666	0.652	0.618	0.683	**0.709**	0.503
Headwear	0.608	0.586	*0.606*	0.584	0.582	0.590	0.606	0.590	**0.618**	0.545	0.594	0.258
Scarf & Tie	0.393	*0.429*	0.396	0.409	0.439	0.426	0.397	0.395	0.420	0.370	**0.473**	0.129
mIoU	0.648	**0.652**	0.643	0.641	*0.650*	0.650	0.647	0.642	0.641	0.619	0.648	0.441
inference time (ms)	100.96	78.99	68.81	68.80	75.11	75.23	*66.16*	66.05	116.24	137.83	115.98	**29.61**
training time (h)	19.568	19.328	17.825	*17.625*	19.359	19.297	17.831	17.483	30.84	33.64	23.68	**9.04**

[a]DeepLabv3+

pixels belongs to the i-th class in j-th ground truth, L_{ij} is the set of all pixels predicted as i-th class in the j-th prediction, and $|\cdot|$ is the cardinality of a set. Thus, the mIoU metric is calculated as follows.

$$mIoU = \frac{1}{M} \sum_{i=0}^{M-1} IoU_i \tag{10}$$

Where M is the total number of segmentation classes.

2) mIoU+: Because the mIoU metric favors the total number of accurately classified pixels, a prediction with noise frequently results in higher mIoU compared to a prediction with no noise. Depending on the application, prediction with low noise may be favored over absolutely high mIoU prediction.

Therefore, we propose mIoU+ (mIoU-plus) metric in which noise is taken into account. This metric is not based on individual pixel but connected components (i.e. individual segments). Given a prediction and a ground truth segmentation, the connected component based segmentation score of a class is calculated as follows.

$$CCSS_i(U, V) = \frac{1}{|U_i|} \sum_{u \in U_i} \max_{v \in V_i} IoU\,(u, v) \tag{11}$$

Where $CCSS_i$ is the segmentation score of the i-th class between predicted segmentation U and ground truth V, U_i is the set of all connected components of i-th class in the prediction, V_i is the set of all connected components of i-th class in the ground truth. However, because this score is not symmetric (i.e., $CCSS_i(U, V) \neq CCSS_i(V, U)$), the segmentation score is calculated as follows.

$$SS_i(U, V) = CCSS_i(U, V) \wedge CCSS_i(V, U) \tag{12}$$

Where $SS_i(U, V)$ is the segmentation score of the i-th class between two segmentations U and V. Similar to the conventional mIoU, IoU+ of each class is computed as follows.

$$IoU{+}_i = \frac{1}{N} \sum_{j=0}^{N-1} SS_i(U_j, V_j) \tag{13}$$

Where $IoU{+}_i$ is IoU+ score of the i-th class, N is the total number of samples, U_j is the set of connected components from the j-th prediction, and V_j is the set of connected components from the j-th ground truth. The score for a whole segmentation with multiple connected components is calculated as follows.

$$mIoU{+} = \frac{1}{K} \sum_{i=0}^{K-1} SS_i(U, V) \tag{14}$$

Where K is the total number of segmentation classes.

D. Ablation Study on Effect of Auxiliary Training Objectives

We investigate more into the effect of auxiliary training objective on the model performance. We retrain our network with different training objective configurations. The loss function used in this experiment is as follows.

$$loss = \frac{H\,(t_0, l_0) + \alpha IPL + \beta SPL + \gamma LPL}{1 + \alpha + \beta + \gamma} \tag{15}$$

Where $\alpha, \beta, \gamma \in \{0, 1\}$. In practice, when $\alpha = 0$, b_i computations are ignored. Similarly, when $\beta = 0$, $l_{[1..4]}$ computations are ignored. However, when $\gamma = 0$, d_i still need to be computed because it is an integrated part of the model.

There are 8 different configurations of the loss function. We annotate them as configuration A to configuration H, as shown in Table V.

TABLE IV
IoU+ OF INDIVIDUAL CLASSES

	A	B	C	D	E	F	G	H	Unet	DLv3+[a]	PSPNet	SegNet
Background	*0.373*	0.344	0.369	0.336	0.359	0.347	0.339	0.338	0.345	0.379	**0.404**	0.233
Bag	**0.441**	0.398	0.434	0.361	*0.429*	0.392	0.424	0.385	0.373	0.364	0.427	0.124
Belt	0.290	0.257	0.312	0.245	**0.316**	0.257	*0.297*	0.240	0.287	0.204	0.211	0.082
Boots	*0.347*	0.303	0.358	0.269	0.335	0.297	**0.360**	0.280	0.297	0.299	0.301	0.100
Footwear	0.502	0.474	0.503	0.483	*0.503*	0.489	**0.504**	0.475	0.500	0.434	0.418	0.247
Outer	0.412	0.379	0.383	0.356	*0.409*	0.375	0.408	0.329	0.286	0.381	**0.425**	0.104
Dress	0.425	0.393	0.370	0.336	*0.398*	0.382	0.393	0.360	0.208	0.366	**0.434**	0.082
Sunglasses	0.501	0.471	**0.541**	0.433	0.531	0.463	*0.514*	0.402	0.502	0.355	0.392	0.157
Pants	0.629	0.613	**0.642**	0.568	*0.635*	0.621	0.623	0.567	0.557	0.591	0.641	0.316
Top	0.432	*0.439*	0.438	0.394	0.452	0.428	0.401	0.391	0.347	0.406	**0.465**	0.173
Shorts	0.488	0.426	*0.490*	0.393	**0.527**	0.454	0.449	0.374	0.364	0.453	0.497	0.203
Skirts	*0.459*	0.413	0.438	0.370	0.465	0.426	0.454	0.390	0.297	0.436	**0.497**	0.163
Headwear	**0.479**	0.399	0.468	0.362	0.452	0.379	*0.467*	0.310	0.441	0.336	0.400	0.083
Scarf & Tie	**0.278**	0.230	0.239	0.185	0.270	0.186	0.238	0.172	0.184	0.211	*0.263*	0.049
mIoU+	0.433	0.396	*0.428*	0.363	**0.434**	0.392	0.419	0.358	0.356	0.373	0.412	0.151
inference time (ms)	100.96	78.99	68.81	68.80	75.11	75.23	*66.16*	66.05	116.24	137.83	115.98	**29.61**
training time (h)	19.568	19.328	17.825	*17.625*	19.359	19.297	17.831	17.483	30.84	33.64	23.68	**9.04**

[a]DeepLabv3+

E. Settings

We implement all the networks using Chainer deep learning framework [27]. LeCun normal weight initializer [28] is used. The models are trained by Adam optimizer [29] with the learning rate of 1×10^{-3} and the decay rate of 0.99. The machine used to carry out the experiment is a Linux box equipped with three Nvidia Pascal GPUs.

We train each configuration for three times with 100 epochs each and take averages of the best mIoU and mIoU+.

F. Result

We report the experimental results using mIoU metric in Table III. Our proposed network configured with three different auxiliary losses outperforms all the ever-proposed models in terms of performance. Among all the auxiliary loss configurations, configuration B achieves the highest mIoU score. This configuration consists of only image pyramid loss and segmentation pyramid loss. Among the ever-proposed networks, PSPNet achieves the highest mIoU score. Moreover, our top proposed network takes only $2/3$ of the time for training as well as inference compared to PSPNet.

We realize that the mIoU performance of the model is worsened when combining label pooling loss with the other two auxiliary training losses. In fact, configuration B, D, and F achieve higher mIoU compared to configuration A, C, and E.

TABLE V
DIFFERENT AUXILIARY CONFIGURATIONS

	A	B	C	D	E	F	G	H
α	1	1	1	1	0	0	0	0
β	1	1	0	0	1	1	0	0
γ	1	0	1	0	1	0	1	0

We report experimental results using mIoU+ metric in Table IV. The training and inference time in Table IV are carried over from Table III. We observe that all the models achieve their best mIoU and mIoU+ in the same epoch. Furthermore, mIoU and mIoU+ are loosely proportional to each other during the training process.

As shown in Table IV, the proposed model with configuration E achieves the highest mIoU+ score. Configuration E consists of segmentation pyramid loss and label pooling loss. Configuration A with all the auxiliary loss functions is the runner-up. Among the ever-proposed model, PSPNet also achieves the highest mIoU+ score.

The proposed model used to generate samples in Fig. 4 and Fig. 5 is trained with configuration A.

V. CONCLUSION

In this paper, we propose a high-performance semantic segmentation model for street fashion photos. This network infers the existence of the classes over the whole image, and progressively refines the result up to the desired resolution. We also propose a new label pool feature that can be used to improve the performance of the proposed network. And finally, we provide benchmarking results of the network with and without 3 different types of auxiliary loss. For better evaluation, we propose mIoU+ metric in which noises are taken into account.

We compare the performance of our network to the other state-of-the-art networks, including U-Net [18], PSPNet [4], SegNet [2], and DeepLabv3+ [3]. We report the evaluation result using both conventional mIoU and newly proposed mIoU+ metrics. The experiment shows that our network requires less time to train and infer while achieves the highest segmentation performance in both mIoU and mIoU+ metrics.

For future work, we will extend the evaluation on the scene parsing problem using data sets such as MSCOCO [5], CityScapes [6], and ADE20K [7].

REFERENCES

[1] J. Long, E. Shelhamer, and T. Darrell, "Fully convolutional networks for semantic segmentation," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2015, pp. 3431–3440.

[2] V. Badrinarayanan, A. Kendall, and R. Cipolla, "Segnet: A deep convolutional encoder-decoder architecture for image segmentation," *IEEE transactions on pattern analysis and machine intelligence*, vol. 39, no. 12, pp. 2481–2495, 2017.

[3] L.-C. Chen, Y. Zhu, G. Papandreou, F. Schroff, and H. Adam, "Encoder-decoder with atrous separable convolution for semantic image segmentation," in *Proceedings of the European Conference on Computer Vision (ECCV)*, 2018, pp. 801–818.

[4] H. Zhao, J. Shi, X. Qi, X. Wang, and J. Jia, "Pyramid scene parsing network," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2017, pp. 2881–2890.

[5] T.-Y. Lin, M. Maire, S. Belongie, J. Hays, P. Perona, D. Ramanan, P. Dollár, and C. L. Zitnick, "Microsoft coco: Common objects in context," in *European conference on computer vision*. Springer, 2014, pp. 740–755.

[6] M. Cordts, M. Omran, S. Ramos, T. Rehfeld, M. Enzweiler, R. Benenson, U. Franke, S. Roth, and B. Schiele, "The cityscapes dataset for semantic urban scene understanding," in *Proc. of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2016.

[7] B. Zhou, H. Zhao, X. Puig, S. Fidler, A. Barriuso, and A. Torralba, "Scene parsing through ade20k dataset," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2017.

[8] S. Zheng, F. Yang, M. H. Kiapour, and R. Piramuthu, "Modanet: A large-scale street fashion dataset with polygon annotations," *arXiv preprint arXiv:1807.01394*, 2018.

[9] J. Shotton, M. Johnson, and R. Cipolla, "Semantic texton forests for image categorization and segmentation," in *2008 IEEE Conference on Computer Vision and Pattern Recognition*. IEEE, 2008, pp. 1–8.

[10] J. Shotton, A. W. Fitzgibbon, M. Cook, T. Sharp, M. Finocchio, R. Moore, A. Kipman, and A. Blake, "Real-time human pose recognition in parts from single depth images." in *Cvpr*, vol. 2, 2011, p. 3.

[11] F. Yu and V. Koltun, "Multi-scale context aggregation by dilated convolutions," *arXiv preprint arXiv:1511.07122*, 2015.

[12] G. Lin, A. Milan, C. Shen, and I. Reid, "Refinenet: Multi-path refinement networks for high-resolution semantic segmentation," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2017, pp. 1925–1934.

[13] L.-C. Chen, G. Papandreou, I. Kokkinos, K. Murphy, and A. L. Yuille, "Semantic image segmentation with deep convolutional nets and fully connected crfs," *arXiv preprint arXiv:1412.7062*, 2014.

[14] ——, "Deeplab: Semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected crfs," *IEEE transactions on pattern analysis and machine intelligence*, vol. 40, no. 4, pp. 834–848, 2017.

[15] C. Szegedy, W. Liu, Y. Jia, P. Sermanet, S. Reed, D. Anguelov, D. Erhan, V. Vanhoucke, and A. Rabinovich, "Going deeper with convolutions," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2015, pp. 1–9.

[16] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," *arXiv preprint arXiv:1409.1556*, 2014.

[17] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2016, pp. 770–778.

[18] O. Ronneberger, P. Fischer, and T. Brox, "U-net: Convolutional networks for biomedical image segmentation," in *International Conference on Medical image computing and computer-assisted intervention*. Springer, 2015, pp. 234–241.

[19] S. Hochreiter, "The vanishing gradient problem during learning recurrent neural nets and problem solutions," *International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems*, vol. 6, no. 02, pp. 107–116, 1998.

[20] T. Chen, X. Zhai, M. Ritter, M. Lucic, and N. Houlsby, "Self-supervised gans via auxiliary rotation loss," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2019, pp. 12 154–12 163.

[21] K. He, X. Zhang, S. Ren, and J. Sun, "Identity mappings in deep residual networks," in *European conference on computer vision*. Springer, 2016, pp. 630–645.

[22] G. Huang, Z. Liu, L. Van Der Maaten, and K. Q. Weinberger, "Densely connected convolutional networks," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2017, pp. 4700–4708.

[23] M. Drozdzal, E. Vorontsov, G. Chartrand, S. Kadoury, and C. Pal, "The importance of skip connections in biomedical image segmentation," in *Deep Learning and Data Labeling for Medical Applications*. Springer, 2016, pp. 179–187.

[24] S. Jégou, M. Drozdzal, D. Vazquez, A. Romero, and Y. Bengio, "The one hundred layers tiramisu: Fully convolutional densenets for semantic segmentation," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops*, 2017, pp. 11–19.

[25] H. Zhao, X. Qi, X. Shen, J. Shi, and J. Jia, "Icnet for real-time semantic segmentation on high-resolution images," in *Proceedings of the European Conference on Computer Vision (ECCV)*, 2018, pp. 405–420.

[26] J. Li, W. Speier, K. C. Ho, K. V. Sarma, A. Gertych, B. S. Knudsen, and C. W. Arnold, "An em-based semi-supervised deep learning approach for semantic segmentation of histopathological images from radical prostatectomies," *Computerized Medical Imaging and Graphics*, vol. 69, pp. 125–133, 2018.

[27] S. Tokui, K. Oono, S. Hido, and J. Clayton, "Chainer: a next-generation open source framework for deep learning," in *Proceedings of workshop on machine learning systems (LearningSys) in the twenty-ninth annual conference on neural information processing systems (NIPS)*, vol. 5, 2015, pp. 1–6.

[28] Y. A. LeCun, L. Bottou, G. B. Orr, and K.-R. Müller, "Efficient backprop," in *Neural networks: Tricks of the trade*. Springer, 2012, pp. 9–48.

[29] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," *arXiv preprint arXiv:1412.6980*, 2014.

Anh H. Dang (S'09) received his bachelor degree in business administration, information & communication technology from Ritsumeikan Asia Pacific University (Beppu, Oita, Japan) in 2010. He then received the master degree in computer science from Waseda University (Shinjuku, Tokyo, Japan) in 2012. Since 2012, he is a Ph.D. candidate at Waseda University. He is a member of IEEE, ACM, and IEICE. His research interests are machine learning, artificial intelligence, and computer vision.

Prof. Wataru Kameyama (M'86) received the bachelor's, master's, and D.Eng. degrees from the School of Science and Engineering, Waseda University, in 1985, 1987, and 1990, respectively. He joined ASCII Corporation in 1992, and was transferred to France Telecom CCETT from 1994 to 1996 for his secondment. After joining Waseda University as an Associate Professor in 1999, he has been a Professor with the Department of Communications and Computer Engineering, School of Fundamental Science and Engineering, Waseda University, since 2014. He has been involved in MPEG, MHEG, DAVIC, and the TV-Anytime Forum activities. He was a Chairman of ISO/IEC JTC1/SC29/WG12, and a Secretariat and Vice Chairman of the TV-Anytime Forum. He is a member of IEEE, IEICE, IPSJ, ITE, IIEEJ, and ACM. He received the Best Paper Award of Niwa-Takayanagi in 2006, the Best Author Award of Niwa-Takayanagi in 2009 from the Institute of Image Information and Television Engineers, and the International Cooperation Award from the ITU Association of Japan in 2012.

APPENDIX A
SEGMENTATION RESULTS

Fig. 5. Illustration of segmentation results coming from the models. Photos are public domain works downloaded from Pexels.com. Label are manually annotated by the authors.

Wearable Passive biosensing interface method for gathering bioinformation

Tae-gyu Lee *

* *Smart Contents Major, Division of* ICT Convergence, *Pyeongtaek University, 3825 Sedong-daero Pyeongtaek-si, Gyeonggi,17869, Korea*

tglee@ptu.ac.kr

***Abstract*— Recently, a variety of biometric information systems have been tried and developed to provide IoT-based healthcare and medical information services. Existing biometric information systems actively collect and analyze biometric information around user terminals or smart devices. On the contrary, this study proposes a passive biosensing system and a passive sensing method for collecting and analyzing biometric information using wearable passive patches. In particular, a passive sensing system and method using a passive patch that can be detected by activation of an external power source without an internal battery is proposed. This paper focuses on the following elements of passive biometric information system. The first describes the structure and components of a passive patch system that can be operated from an external source without an internal battery. The second describes passive biometric information algorithms that passively detect biometric information in real time. Finally, we propose a passive sensing analysis modeling that analyzes and evaluates the signal sensitivity of biometric information, and describes an analysis example of bioinformation sensitivity.**

***Keyword*—Passive sensing, Patch interface, IoT, Wearable system, Bioinformation**

I. INTRODUCTION

WITH the development of various wearable systems, the methods for monitoring a human body, analyzing a bio-signal, evaluating a health condition of a user, and providing a healthcare service based on the evaluation have been continuously attempted. These efforts have had some fruitful consequences, and have become a future vision strategy that is important for the healthcare sector. This paper describes a sensing system and a sensing method based on bio-information collection interface using passive patches. Especially, it is a research on sensing system and sensing method based on information gathering interface using passive patch which can be sensed by the activation of external power without internal battery [1].

When a drug is administered to treat a patient, it is largely divided into oral administration and parenteral administration. In this case, the parenteral administration may be divided into mucosal administration methods such as ocular mucosal administration, nasal administration, sublingual administration and rectal administration, and injection administration methods such as intravenous injection and subcutaneous injection, and skin administration method [2-6]. Among them, the method of administering the drug through the skin is as shown in Fig. 1, using a paste, a patch, or a wetting agent.

Until now, digital patches have been used as means and methods for collecting human body information. These efforts contributed to the collection of various biometric information such as body temperature, sweat, humidity, respiration, pulse, and heart rate. Nevertheless, the types of sensors available for patching are limited and the amount of information that can be collected is limited. These attempts should continue to develop new sensors, patch structures, and bio-signal analysis.

This paper considers the following issues in order to realize a passive patch system that maximizes applicability. First, there is a need for an alternative to overcome the mobile support and user convenience constraints of a typical power built-in active sensing patch. Next, a power supply alternative is needed with a no-power biometric information system configuration for semi-permanent product and service configuration. Finally, by providing a consumable patch sensing module according to the no-power patch system configuration, the patch production cost is minimized to maximize the applicability.

The subjects of this study are as follows [7]. The first is to provide an organization of a passive patch system based on an bioinformation gathering interface that can operate with external power without an internal battery. The second is to provide a passive patch based on an information collection interface that can sense the degree of action of the applied drug and display it to the patient or the user. The third is to provide a passive patch based on an information gathering interface that can inform the user or the patient when the degree of pharmacological action is out of the threshold and can lead the user or the patient to remove the patch attached. Finally, this study is to provide a sensing system and a sensing method using a passive patch based on an information collection interface capable of acquiring biometric information and providing it to an external network.

This study aims to classify and define biometric information sensing method in systematically developing and implementing biometric information sensing and biometric information system [8] [9] [10].

Manuscript received on June 21, 2019. This work is a follow-up the invited journal to the accepted & presented paper of the 20th International Conference on Advanced Communication Technology (ICACT2018), and Grant ID is 2020001.

T. L. Author is with the Division of ICT Convergence (Smart Contents Major), Pyeongtaek University, Pyeongtaek, Gyeonggi 17869 Korea (corresponding author to provide phone: +82-31-659-8370; fax: +82-31-659-8011; e-mail: tglee@ptu.ac.kr).

First, there is an *active bio-sensing method* for sensing and collecting biometric information of a target object in accordance with its own bio-logic execution while continuously supplying power and electric power to the sensing module.

Second, the sensing module can be configured with passive sensing logic without its own built-in power or without external power supply. In order to sense and collect biometric information of a target object, a sensing module is transferred to a wireless local power and activation (or synchronization) signal, and necessary biometric information is sensed through a sensing circuit and a wireless transmission circuit module that are passively operated It is a *passive bio-sensing method* to collect.

This study applies passive bio-sensing method without built-in power in the patch module as bio-information sensing module and system configuration method. The main issues to build processes and systems for collecting, storing, and processing biometric information are as follows. An active agent and a wireless interworking interface configuration for activating the passive sensing module are needed. In particular, consider how to design the signal activation, the available frequency band and the signal strength. As a layered structure of the passive sensing module, process support such as activation signal reception, individual bio-information sensing, biometric information transmission and power signal amplification are important issues. Finally, the real-time signal analysis capability of the active agent and the non-real-time bio-information application data model of the background server are recognized as major issues.

The application fields of biometric information technology and system of this study are as follows. Currently, it is active in the medical and healthcare fields that utilize mobile platforms such as smart phones and smart clocks [11]. It has been steadily expanding in the fields of sports and clothing [12], beauty sector [13], agricultural biology sector [14], automobile and transportation sector [15] and security certification sector [16].

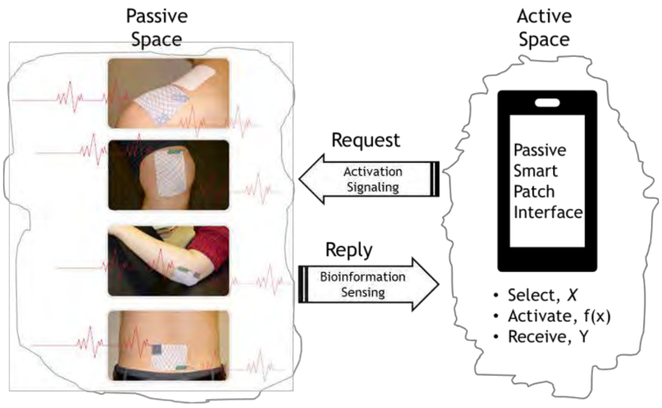

Fig. 1. A Computing Domain Configuration for Wearable Passive Patch Bioinformation System

This study has the following contributions [7]. The passive patch based on the proposed biometric information collection interface can operate with external power without built-in battery, it is easy to miniaturize, shorten manufacturing process time, and reduce production cost. And, such a passive patch can measure the degree of action of the drug and induce removal of the patch when out of the threshold range. Additionally, it can transmit the biometric information of the patient to the outside and transmit the action information of the internal medicine of the patch used for the treatment to the outside, so that the transmitted data can be easily browsed. Accordingly, the patient, the caregiver or physician can conveniently check the patient status, and can transmit a patch removal message or the like when the medicinal threshold of the passive patch attached for the patient treatment is out of the range. Finally, the sensing system using the passive patch based on the proposed bio-information collecting interface receives the biometric information transmitted from the passive patch and transmits the biometric information to the registered communication terminal when the biometric information is out of the threshold range.

II. Architecture Of Passive Sensing System

A. *Component and functions*

In this study, a passive patch attached to a patient and measuring the biometric information signal of a patient and transmitting it to the outside, and an active terminal connected to the passive patch, transmits an activation signal to the passive patch, and receives the biometric information signal as shown in Fig. 2 [7]. This passive patch constitutes an information collecting interface based passive patch sensing system characterized in that external power is temporarily supplied without operation of an internal battery.

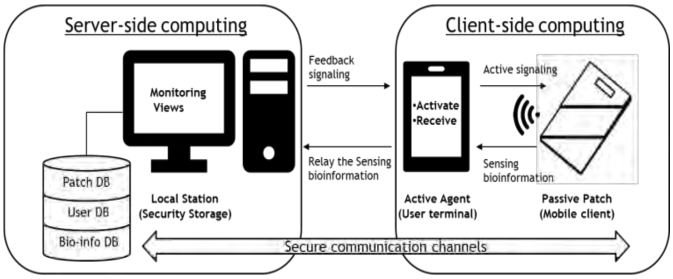

Fig. 2. The Wearable Bioinformation Computing System with a Passive Patch Interface

In addition, the expansion system may further include an external server connected to the active terminal in a wireless or wired manner, and the external server may include a passive information collection interface-based patch sensing system capable of receiving a biometric information signal of a patient from an active terminal acting as a mobile agent.

Here, as shown in Fig. 3, the passive patch constitutes several layers of partial modules from the bottom to the top as follows. The first lower layer constitutes the drug application layer on which the drug is formed on the lower surface which is in close contact with the skin of the patient. The second layer is a supporting module formed on the upper part of the lower drug coating layer, and includes a biosensor module for measuring a biometric information signal of a patient and a drug action detection sensor module for generating a drug action signal by measuring the degree of the drug acting on the patient do. And a power supply module for operating the built-in circuit by receiving power from the outside. The third layer includes a display module formed on the upper part and a control module for generating a patch removal inducing signal corresponding to the drug operation signal when the drug action signal is out of a predetermined threshold value range and transmitting the signal to the display module.

The sensing data transmission structure of the passive patch system is as follows. First, the active terminal performs an activation process to send an activation signal to the passive patch attached to the patient. Second, a passive patch that transmits an activation signal from the active terminal proceeds with a biological information sensing step of sensing drug and human body information. Third, the passive patch consists of a feedback step of feeding the sensed drug and human body information to the active terminal. Finally, the active terminal constructs a storage process for storing the sensed drug and human body information in the internal memory. At this time, the passive patch can be operated by temporarily supplying external power without an internal battery.

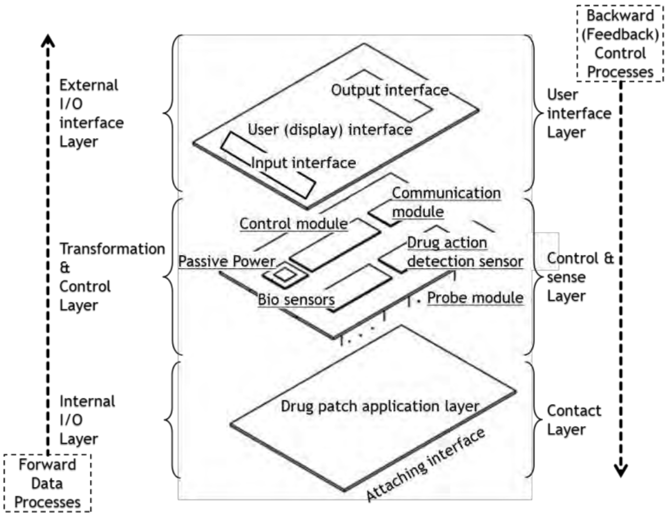

Fig. 3. The Layered Architecture and Components of Passive Patch on the Client-side

Further, the drug information and the human body information stored in the active terminal can be further expanded to an external transfer process for transferring and storing the drug and the human body information to a database of the external server.

B. Data and control flows

Data flow from data collection to data analysis and evaluation service of bio-information sensing system is shown in Fig. 4.

The data flow of the passive biometric information system starts data input and sensing at the passive object by the activation signal of the active agent.

Then, data processing such as data conversion, data transmission and storage based on data compression, real-time data pattern and semantic analysis, non-real-time background data analysis, and data generation of biometric information application services are performed.

The passive-based sensing data flow mainly collects a small amount of data on the smart agent on a non-periodic basis, and provides a real-time alarm and a status notification service of the biometric information to the user or the terminal according to the occurrence of the event.

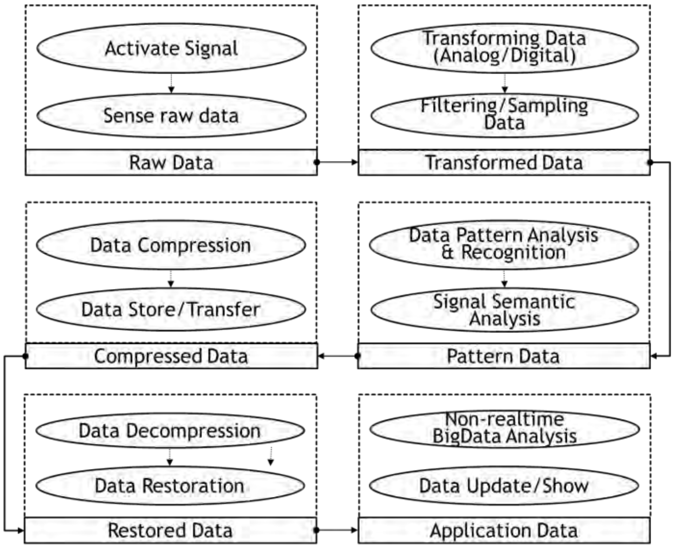

Fig. 4. Data Flow Processes on Bioinformation System

In addition, the information collected in the smart agent terminal is transmitted to a PC or a server located in the background, continuously analyzed and evaluated, and updates various bio-information.

III. Algorithm for Passive Sensing Processes

The detailed components for implementing the passive patch biometric information system include a drug action detection sensing module, a biosensor module, a power supply module, a control module, a communication module, and a display module as follows.

The drug action detection sensing module can detect the drug application layer, or the concentration of the drug detected in the skin of the patient. Further, the drug action detection sensing module may generate a drug action signal through the detected signal and transmit the drug action signal to the control module. Here, the drug action detection sensing module may be formed through at least one probe to contact the drug application layer or the skin of the patient.

The biosensor module can measure a bio-information signal including at least one of body temperature, pulse, blood sugar, skin humidity, electrocardiogram signal, and brain wave. In addition, since the biometric information signal measured according to the attachment position of the passive patch differs, the bio-sensing module can select a different type of sensor depending on the attachment position.

The power supply module may be formed to be able to receive power temporarily from the outside. To this end, the power supply module may configure a USB terminal or configure an electric wire or a pin to receive external power. At this time, the power supply module may be connected to the built-in circuit formed in the control module. That is, the power supply module can receive the power temporarily from the outside to operate the built-in circuit, whereby the passive patch can be operated. Meanwhile, the power supply module may receive power wirelessly. At this time, the wireless power supply may be an electromagnetic induction type, the power supply module may constitute a primary coil, and electromagnetic can be guided through a wireless supply device having a secondary coil to supply power to an internal circuit.

The control module may receive the drug action signal input from the drug action detection sensing module and compare the received drug action signal with a pre-set threshold value. Also, the control module may generate a patch removal inducing signal and display the patch removal inducing signal on the display module when the value included in the drug action signal is out of the threshold value range. For example, if the concentration of the drug contained in the received drug action signal is lower than the threshold value, the control module can generate a notification that the patient can remove the passive patch, that is, a patch removal inducing signal because the drug treatment effect is low. In addition, the control module can generate a patch removal inducing signal so that the patient removes the passive patch because the drug may cause side effects when the concentration of the drug contained in the received drug action signal is higher than the threshold value.

The communication module may transmit a biometric information signal or a patch removal inducing signal or a drug action signal to the outside of the passive patch. Further, the communication module may include wireless communication means capable of communicating at a short distance, such as Bluetooth. For example, the communication module can transmit a biometric information signal or a patch removal inducing signal or a pharmacological action signal to a nearby electronic communication terminal (for example, a smart phone) in a Bluetooth manner. At this time, the communication module may use a module to which the ZigBee communication method or the NFC (Near Field Communication) communication method other than the blue pitching method is applied.

The display module may be formed on the upper surface to display a signal informing of the removal of the patch. The display module also receives the patch removal induction signal and displays the set colour or message to inform the patient or user of the patch removal.

As described above, the passive patch described in this study can be powered by external power supply without built-in battery.

In addition, biometric information of the patient can be transmitted to the outside, and action information of the internal medicine of the patch used for the treatment can be transmitted to the outside, so that the transmitted data can be easily browsed. In this study, an active terminal as an agent terminal can use a WIFI or Bluetooth network or the like, and thus can connect to an external server using such a communication network.

In addition, the active terminal and the external server can store the subscribed patient information. At this time, the patient information may be stored through one of the active terminal or the external server and shared with each other. In other words, when stored in the active terminal, the stored information is transmitted to the external server and stored in the external server, transferred from the external server to the active terminal upon storage, and transmitted to the active terminal.

Here, the reference health information may be reference health information classified according to the patient's age, sex, body shape or the like or pre-set reference health information (for example, blood pressure, pulse, blood glucose level, etc.).

This standard health information can be used by backing up the data stored in the existing hospital database for the recent health condition information and medical history information. If necessary, the patient can receive the data from the treated hospitals and store it on the active terminal or the external server is.

This is an extended feature of the sensing system using the passive patch based on the information gathering interface so that it is possible to manage the health and medical information according to each user, thereby enabling more efficient additional service support.

The proposed passive biometric information sensing system is divided into two parts. One is the passive patch portion of Fig. 5, and the other is the mobile active agent portion shown in Fig. 6. The sensing module of the passive patch waits for the activation signal from the mobile active agent, and the mobile active agent transmits the activation signal to the passive patch whenever biometric information collection is required. The details of each process are shown below.

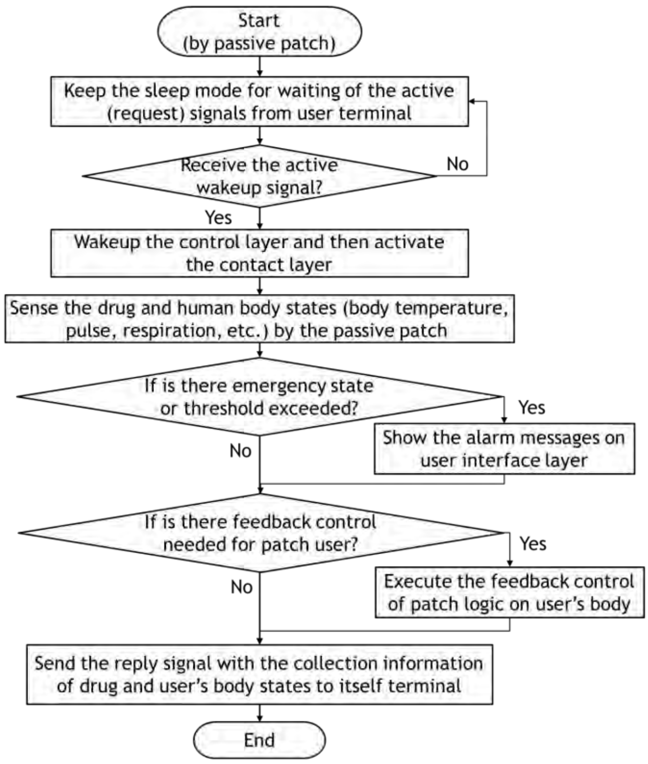

Fig. 5. The Embedded Logic Sequences of Passive Patch

Fig. 5 is a flow chart showing a passive sensing method using passive patches based on the bio-information collection interface. The bio-sensing process using the passive patch is as follows. First, a process in which the active terminal of a client transmits an activation signal to a passive patch attached to the patient. Then, a passive patch that receives an activation signal from an active terminal and wakeups itself senses drug or human body bio-information such as body temperature, pulse, respiration, etc. At this time, if the bio-sensing information exceeds the threshold or an emergency occurs, the alert message is notified to the patch display interface. Third, a reply process returns the sensed drug and human body information to the active terminal. At this time, feedback control can be performed on the passive patch if necessary. Finally, the active terminal carries out a process of receiving the sensed drug and the human body information, and storing it therein.

Fig. 6 shows the processes of collecting sensing information received from a user mobile terminal and transmitting information to an external server as a mobile agent.

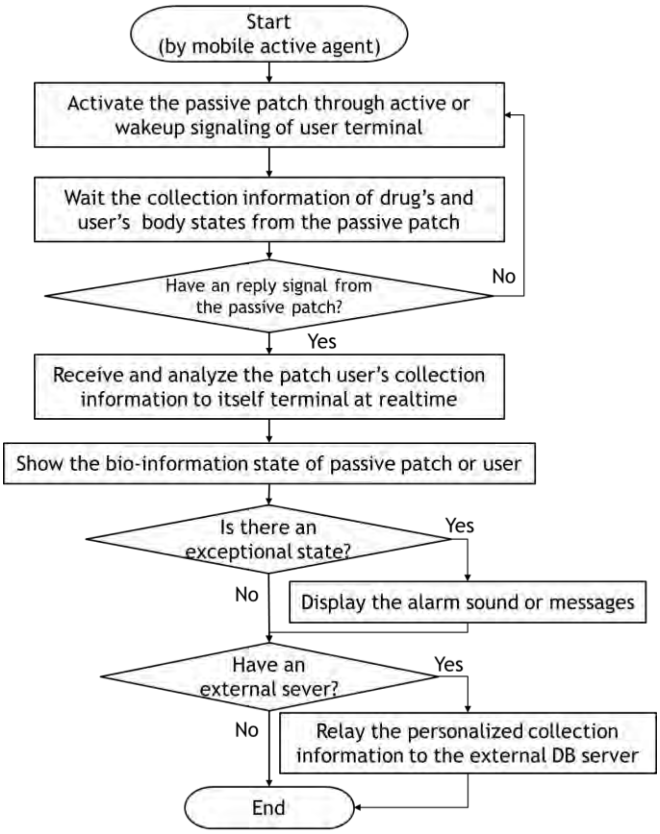

Fig. 6. The Active Agent Logic Sequences of Mobile Terminal

In Fig. 6, the user terminal initially activates the passive patch with an active signal. Next, it waits for receiving the bioinformation and the user's biometric information from the passive patch. If a reply message is detected, it stores the received bio-information in its storage space, and displays the bio-information state of passive patch at real-time. Then, if there is an exceptional state, it can show the alarm services on the mobile terminal. Finally, when the external DB server is existed and interconnected, the information collected through the external communication channel can be relayed and constructed as a database.

IV. ANALYSIS MODEL OF PASSIVE SENSING

This study first proposed a passive bioinformatics system. And, this configuration can be a good alternative to overcome the power consumption limitation of the digital patch. In addition, the cost of producing a digital patch is minimized, thereby maximizing the economic efficiency of the product. Ultimately, we will expect the availability of passive digital patches to vary.

A. *Experimental environments*

Passive RFID systems use tags with no internal power source and instead are powered by the electromagnetic energy transmitted from an RFID reader. The wireless communication configuration of the passive sensing data transmission system for passive patch sensing is as follows.

As the name implies, passive tags wait for a signal from an RFID reader. The reader sends energy to an antenna which converts that energy into an RF wave that is sent into the read zone. Once the tag is read within the read zone, the RFID tag's internal antenna draws in energy from the RF waves. The energy moves from the tag's antenna to the IC and powers the chip which generates a signal back to the RF system. This is called backscatter. The backscatter, or change in the electromagnetic or RF wave, is detected by the reader (with the antenna), which interprets the information.

Combining sensor monitoring with RFID allows for the observation of uniquely identifiable items at short and long ranges. Whether using battery power or collected RF energy, sensors can collect relevant data pertaining to temperature, humidity, moisture, and motion/movement of human body. With advancements in technology, especially in energy collection and impedance, sensor monitoring can now be accomplished using passive RFID tags without a decrease in read range. Recently, manufacturers have figured out a way to incorporate sensors into passive RFID tags without greatly affecting read range. Because passive tags do not have an internal battery to power the sensors, passive tags use RF energy to power the IC/sensor and to send the relevant information back to the reader. Currently, passive RFID sensor tags can be used to detect body states such as temperature and humidity or moisture because of a newly wearable integrated circuit (IC). The IC is designed with a bank of capacitors that is able to detect the presence of moisture based on how the moisture affects the tuning of the antenna. That information is then backscattered back to the RFID reader for observation.

RFID systems, readers and tags communicate mostly through the method of electromagnetic coupling. In order for an RFID tag to communicate with an RFID reader/antenna, the tag circuit and reader circuit generally must couple in some way. Coupling is a transfer of energy between two electronic items or two circuits. For example, systems that use capacitive coupling use electric currents instead of the magnetic field in order to couple. Contact to a few centimetres of read range is normal for (LF) or Low Frequency communication because of the need to produce an electric field using electrodes.

Passive RFID tags do not all operate at the same frequency. There are three main frequencies within which passive RFID tags operate. The frequency range, along with other factors, strongly determines the read range, attachment materials, and application options.

As the first frequency model, Low Frequency (LF, 125~ 134 KHz) has an extremely long wavelength with usually a short read range of about 1~10 centimetres. This frequency is typically used with animal tracking because it is not affected much by water or metal.

The second frequency model with High Frequency (HF, 13.56MHz) & Near-Field Communication (NFC) has a medium wavelength with a typical read range of about 1 centimetre up to 1 meter. This frequency is used with data transmissions, access control applications, DVD kiosks, and passport security – applications that do not require a long read range.

The third model with Ultra High Frequency (UHF, 865~960 MHz) has a short, high-energy wavelength of about a one meter which translates to long read range. Passive UHF tags can be read from an average distance of about 5 – 6 meters, but larger UHF tags can achieve up to 30+ meters of read range in ideal conditions. This frequency is typically used with race timing, IT asset tracking, file tracking, and laundry management as all these applications typically need more than a meter of read range.

As a general rule, higher frequencies will have shorter, higher-energy wavelengths and, in turn, longer read ranges. Moreover, the higher the frequency, generally speaking, the more issues an RFID system will have around non-RFID-friendly materials like water and metal.

B. *Analysis and evaluation Model*

This study starts by collecting human body and patch application data based on a sensing patch module by constructing a human - based wearable patch.

The factors affecting the wearable sensing system are sensing sensitivity and sensing success rate, sensing data rate, power consumption, production and component cost, sensing scalability, mobility, lightweightness, and many other environmental variables.

Nevertheless, this study focuses primarily on the sensing sensitivity and the ability to transmit sensing data as the main environmental variables for analysing and evaluating the performance of passive sensing patches.

Let us assume that the power transmitted by the reader has a uniform power density in all directions over a spherical surface at any given distance *d*. Some of this power is received by the tag antenna and is proportional to the effective aperture of the tag antenna and the power impinging on the tag [17].

In RFID, the distance between the reader and passive tag has a very significant effect on the power loss which decreases as the inverse square of the distance between the reader and passive tag. The power density received at a distance *d* is given by the free space transmission formula,

$$P_d = \frac{P_r}{4\pi d^2}$$

where P_d is the received power density, P_r is the power radiated by the reader antenna and *d* is the distance between the reader and passive tag.

First, the sensing sensitivity represents the reception sensitivity for each radius according to power and active signal strength as shown in Fig. 7.

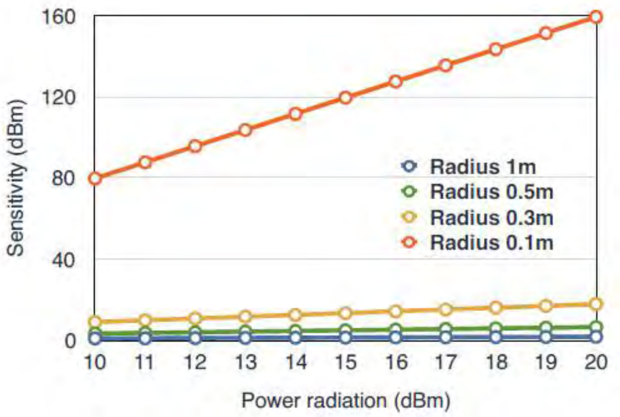

Fig. 7. Passive Signal Sensitivity over Power Radiation

Second, the receiving sensitivity is expressed according to changing in transmitter power of the received signal versus the signal transmission distance between the active terminal and the passive patch as shown in Fig. 8.

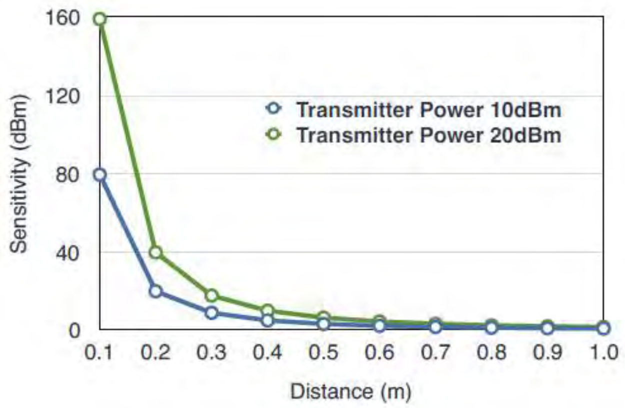

Fig. 8. Receiving Signal Sensitivity over Distances

V. CONCLUSION

In this paper, we first proposed the architecture and components of passive biometric information system. This proposed system can be a good alternative to overcome the limitation of power consumption of digital patches. In addition, the cost of manufacturing a digital patch is minimized, maximizing the economic efficiency of the product and maximizing market usability.

Then we proposed biometric information collection and feedback algorithms, proposed an analytical model of passive sensing system, and described analysis evaluation examples. This suggests a future direction for the design of advanced biometric information system.

In the future, this study intends to expand the scope of research by constructing multi passive patch biometric information system. We intend to build an embedded bio-sensing platform for the development of embedded bio-sensing PCB / Chip.

REFERENCES

[1] T. G. Lee and S. H. Lee, “Dynamic Bio-sensing Process Design in Mobile Wellness Information System for Smart Healthcare s”, *Wireless Personal Communications*, Vol. 86, No. 1, p. 201-215, 2016.

[2] P. Celka, R. Vetter, P. Renevey, C. Verjus, V. Neuman, J. Luprano, J. D. Decotignie and C. Piguet, “Wearable biosensing: signal processing and communication architectures issues”, *Journal of Telecommunications and Information Technology*, p. 90-104, 2005.

[3] T. G. Lee and S. H. Lee, “Design of Wearable Bio-patch System Platform in Human Healthcare Environment”, *Indian Journal of Science and Technology*, Vol. 8, p. 1-7, 2015.

[4] S. Patel, H. Park, P. Bonato, L. Chan and M. Rodgers, “A review of wearable sensors and systems with application in rehabilitation”, *Journal of NeuroEngineering and Rehabilitation*, Vol. 9, No. 21, 2012.

[5] T. G. Lee, M. S. Ko and S. H. Lee, “Wearable Multiple Bio-sensing Process Architecture in Human Healthcare Environments”, *International Journal of Bio-Science and Bio-Technology*, Vol. 6, No. 5, p. 177-184, 2014.

[6] Tae-Gyu Lee, *Chapter 15. Mobile Healthcare Computing in the Cloud, Mobile Networks and Cloud Computing Convergence for Progressive Services and Applications*, IGI Global, p. 275-294, 2014.

[7] T. Lee, "Sensing system and sensing method of passive patch interface for gathering bioinformation," The 20th International Conference on Advanced Communication Technology (ICACT), Chuncheon-si Gangwon-do, Korea (South), 2018, pp. 652-655. doi: 10.23919/ICACT.2018.8323869

[8] Ki-hwan Chon and Hyung-jin Choi, "A Study on Ubiquitous Psychological State Recognition Model Using Bio-Signals," The Journal of Korean Institute of Communications and Information Sciences, Vol. 35, No. 2, pp. 232~243, 2010.

[9] Jin Seob Shin, Woo Young Ahn and Il Yong Oh, "Development of Diagnosis System through Human-body Information Measurement," Journal of the Korea Society of Computer and Information , Vol. 13, No. 1, pp. 219~226, 2008.

[10] Jee Hyun Lee, Eun Ji Lee, Ji Eun Kim, Yoo Lee Kim and Sin Won Cho, "Wearable Sensing Device Design for Biological Monitoring," Journal of the Korean Society of Costume, Vol. 65, No. 1, pp. 118~135, 2015.

[11] Ju-Hyun Choi, Seung-Man Chun, Dong-Hyun Jang and Jong-Tae Park, "Design and Implementation of Bio-data Monitering System Based on ISO/IEEE 11073 DIM/REST for IoT Healthcare Service," Journal of the Institute of Electronics and Information Engineers, Vol. 52, No. 3, pp. 3~12, 2015.

[12] Tae-Gyu Lee, Gi-Soo Chung and Seong-Hoon Lee, "Textile IT Convergence: Digital Clothing and Wearable Computing Status," Korea Institute of Information Technology Magazine, Vol. 10, No. 2, pp. 77~86, 2012.

[13] Eun Hye Park and Seung Jin Moon, "U-Healthcare offers and designed wired and wireless integrated Hair care System," Proceedings of Symposium of the Korean Institute of communications and Information Sciences, pp. 698~699, 2010.

[14] Y. S. Kwon, "Greenhouse environmental management by phytomonitoring," Protected Horticulture and Plant Factory, Vol. 8, No. 1, pp. 127~130, 1995.

[15] Tae-hwan Park, Hwa-joeng Seo, Ji-hwan Lim and Ho-won Kim, "Suggestion of Secure Driver Authentication and Vehicle Control System based on NFC Communication and Biometric Information," Journal of the Korea Institute of Information and Communication Engineering, Vol. 22, No. 4, pp. 700~707, 2018.

[16] Myung Geun Chun and Ki-young Moon, "Use of biometric information and protection of privacy," REVIEW OF KIISC, Vol. 15, No. 6, pp. 11~18, 2005.

[17] R. B. Green, "The general theory of antenna scattering," in Ph.D. Thesis, The Ohio State University, Columbus, OH, 1989.

Tae-Gyu Lee (BSc'92–MSc'96–PhD'06) received the B.Sc. degree from Kunsan National University, Kunsan, Korea in 1992, the M.Sc. degree from Soongsil University, Seoul, Korea in 1996, and the Ph.D. degree from Korea University in 2006. He is currently a Professor in the Smart Contents Major, Division of ICT Convergence, Pyeongtaek University, Gyeonggi, Korea since 2018. He has been a Professor in the Support center for Field Practice Education, Wonkwang University, Jeonbuk, Korea for 2014-2018. He has been a Professional Researcher in Advanced Convergent Technology R&D Group, Korea Institute of Industrial Technology (KITECH), Ansan, Korea for 2009-2013. He has also been a President in the JIGUNET Corporation, Seoul, Korea, from 1999. His research interests are in distributed systems, ubiquitous computing, middleware, networks, bioinformatics, wearable and robot computing. Prof. Lee is currently a lifetime member of Korea Information Processing Society, IEEE ICACT reviewer, IEEE TPDS reviewer.

Volume 8 Issue 6, November. 2019, ISSN: 2288-0003

ICACT-TACT
JOURNAL

GIRI
Global IT Research Institute

1713 Obelisk, 216 Seohyunno, Bundang-gu, Sungnam Kyunggi-do, Republic of Korea 13591
Business Licence Number : 220-82-07506, Contact: tact@icact.org Tel: +82-70-4146-4991