FLDNet: Light Dense CNN for Fingerprint Liveness Detection

YONGLIANG ZHANG1, SHENGYI PAN1, XIAOSI ZHAN2, ZHIWEI LI3, MINGHUA GAO1, AND CHENHAO GAO1

1College of Computer Science and Technology, Zhejiang University of Technology, Hangzhou 310023, China
2School of Science and Technology, Zhejiang International Studies University, Hangzhou 310023, China
3Hangzhou Jinglianwen Technology Company Ltd., Hangzhou 310014, China

Corresponding author: Yongliang Zhang (titanzhang@zjut.edu.cn)

This work was supported in part by the Public Welfare Technology Research Program of Zhejiang Province under Grant LGF18F030008, and in part by the Hangzhou Jinglianwen Technology Company Ltd.

ABSTRACT Fingerprint liveness detection has gained increased attention recently due to the growing threat of spoof presentation attacks. Among the numerous attempts to deal with this problem, the Convolutional Neural Networks (CNN) based methods have shown impressive performance and great potential. However, there is a need for improving the generalization ability and reducing the complexity. Therefore, we propose a lightweight (0.48M parameters) and efficient network architecture, named FLDNet, with an attention pooling layer which overcomes the weakness of Global Average Pooling (GAP) in fingerprint anti-spoofing tasks. FLDNet consists of modified dense blocks which incorporate the residual path. The designed block architecture is compact and effectively boosts the detection accuracy. Experimental results on two datasets, LivDet 2013 and 2015, show the proposed approach achieves state-of-the-art performance in intra-sensor, cross-material and cross-sensor testing scenarios. For example, on LivDet 2015 dataset, FLDNet achieves 1.76% Average Classification Error (ACE) over all sensors and 3.31% against unknown spoof materials compared to 2.82% and 5.45% achieved by state-of-the-art methods.

INDEX TERMS Fingerprint liveness detection, convolutional neural networks, presentation attacks, attention pooling.

I. INTRODUCTION

Currently, Automatic Fingerprint Identification System (AFIS) is widely applied in many day-to-day applications, including unlocking smartphones, mobile payment, personal identity authentication, etc. At the same time, the security of AFIS is of growing concern due to the challenge of fingerprint spoof attacks [1]. Artificial fingerprint replicas, also known as spoofs, can be easily fabricated by various inexpensive and commonly available materials, such as gelatin, silicone, wood glue and play-doh [2], [3]. Besides, sophisticated 3D printing techniques are also utilized in fingerprint spoof attacks [4]. Fingerprint liveness detection (FLD) is regarded as a primary countermeasure against this growing threat, which prevents spoof attacks by analyzing images captured from live fingers or fake ones [5].

Generally, the various proposed anti-spoofing methods can be roughly divided into two categories, hardware-based and software-based approaches [1], [6]. For hardware-based methods, extra sensors are utilized to detect the characteristics of vitality, such as temperature, blood flow, pulse oximetry and odor [7], [8]. Although the hardware-based solutions can prevent the spoof attacks to some extent, they also inevitably increase the cost and the complexity of the whole system. More importantly, these additional hardware devices are difficult to update in time once they are beaten by some new types of spoofs. The software-based solutions, however, distinguish the live and spoof fingerprints by features extracted from the captured fingerprint images. Since no additional hardware cost is incurred, the software-based solutions are more cost-effective and easier to update [9].

The software-based solutions in the literatures are typically based on one of the following features: sweat-pores, perspiration, skin-elasticity, image-quality and texture-feature [10]–[14]. Texture-based solutions, distinguishing the live
and fake fingerprints through the differences in texture features including continuity, clarity and ductility, have attracted researchers’ attention. Local binary pattern (LBP) based on gradient [15] is the first application of texture features in FLD, which compared the value of the central pixel with its adjacent pixels. In literature [16], modifications were made to this method. The proposed uniform local binary pattern greatly improved the performance on standard datasets. Gragnaniello et al. [17] combined the gradient and local phase information together and named it as local contrast phase descriptor (LCPD). Xia et al. [18] proposed a novel weber local binary descriptor (WLBD), consisting of the local binary differential excitation to extract intensity-variance features and the local binary gradient orientation to extract orientation features. Recently, Soler et al. [19] analyzed the combination of Scale Invariant Feature Transform (SIFT) with three different feature encoding approaches, aiming to improve the accuracy of detecting spoofs built from unknown materials or captured by different sensors.

Most of the existing texture-based methods lean on professional domain knowledge to engineer the handcrafted features. However, these features are difficult to generalize due to the lack of robustness against unknown materials and sensor diversity [20]. Moreover, texture descriptors, a kind of shallow feature, only reflect the surface properties of the fingerprint images, while leaving those essential ones unrevealed [21].

In contrast to approaches using traditional handcrafted features, most state-of-the-art methods are learning based, where the high-level semantic features are learned by training Convolutional Neural Networks (CNN). Nogueira et al. [22] introduced a pre-trained VGG model to spoof detection, which significantly outperformed the previous algorithms and won the first place of the Fingerprint Liveness Detection Competition 2015 (LivDet 2015). Jang et al. [23] improved the detection accuracy of CNN by performing contrast enhancement. In literature [24], researchers proposed a metric learning approach based on triplet convolutional networks instead of the traditional binary classification model, where the liveness detection was realized by matching the patches from the test image against a set of reference live and fake patches. Chugh et al. [25] proposed a CNN-based method with a voting strategy on minutiae-centered local patches, which provided remarkable classification accuracy. However, locating a large number of minutiae and taking every patch into account will inevitably increase the computational cost and the processing time, thus making it unsuitable for real-time applications. Nguyen et al. [26] proposed a compact and efficient network architecture consisting of Fire and Gram-K modules, whose performance was comparable to state-of-the-art accuracy while the network size was significantly reduced. Deep Residual Network is firstly applied to FLD in [21], where an adaptive learning mechanism and a region of interest (ROI) extraction algorithm were also put forward. Recently, Zhang et al. [27] modified the original residual network. The new architecture, named Slim-ResCNN, was relatively light-weight yet powerful and won the first prize in LivDet 2017.

However, most proposed deep learning techniques use fingerprint images to fine tune the pre-trained CNN models, such as GoogLeNet [20], VGG-19 [22] and MobileNet-v1 [25], rather than redesigning a network specifically for FLD. The unneglectable differences between fingerprint images and natural images prevent it from achieving high accuracy by simply applying general CNN structures. Moreover, there is a new trend that FLD is gradually moving to the embedded devices and mobile devices where only a limited computation and storage resources are available. However, the method proposed in [25] took 100ms to classify a single input image on Nvidia GTX 1080 Ti GPU. Thus, it is urgent to develop a lightweight deep learning algorithm so that spoof detection can be deployed on low specification systems.

To address the problems mentioned above, a compact and efficient network structure, named FLDNet, is proposed. We firstly discuss the weakness of Global Average Pooling (GAP) in fingerprint anti-spoofing tasks, and overcome it by adopting the attention mechanism. The effectiveness of the proposed attention pooling layer is further verified by comparative experiments. Then, a new block architecture is designed based on the original dense block, where both the residual path and the densely connected path are incorporated, benefiting from two network topologies. Besides, before sending the finger-print images into the network, region of interest (ROI) extraction is firstly applied to eliminate the interference of background areas. Then five local patches are segmented to enlarge the training sets, preventing network from overfitting. A score fusion of these patches are performed during the testing stage. Moreover, except for standard data augmentation methods (rotation and flipping), mixup is also applied in the training to further strengthen the generalization ability. The main contributions of this paper are enumerated as follow:

- A lightweight and efficient CNN architecture (FLDNet) is proposed. The new architecture consists of specially designed blocks where the residual path and the densely connected path are incorporated.
- Discussed the limitation of GAP in FLD tasks and overcame it by adopting the attention mechanism. The proposed attention pooling layer differentiates the importance of units at different positions of the feature map.
- Experiments under three scenarios (intra-sensor, cross-material and cross-sensor) show that the proposed approach outperforms the results published on LivDet 2013 and 2015 datasets. For example, FLDNet achieves 1.76% Average Classification Error (ACE) over all sensors on LivDet 2015 compared to 2.82% achieved by state-of-the-art methods [19].
- FLDNet is suitable for the deployment on low specification systems due to its compactness (0.48M parameters) and low processing time.
The remainder of the paper is structured as follows: the proposed method is discussed in Section II. Section III presents the experimental evaluation. Finally, the conclusions and the directions of future research are given in Section IV.

II. PROPOSED METHOD

In this section, an efficient and lightweight network architecture, named FLDNet, is specially designed for fingerprint liveness detection. The image preprocessing is firstly performed, which consists of region of interest (ROI) extraction and patch segmentation. Inspired by the limitation of Global Average Pooling (GAP) in relevant tasks, the attention mechanism is adopted to modify the original GAP. Furthermore, to achieve higher accuracy with a low computational cost, a new block architecture is introduced.

A. PREPROCESSING

The preprocessing consists of two steps, ROI extraction and patch segmentation. To avoid the impact of the background area on network efficiency and performance, ROI extraction is firstly applied to extract the fingerprint foreground region. The ROI is determined by a combination of local mean of grey-scale and local variance of gradient magnitude of the fingerprint image [28].

After the ROI extraction, local patch with fixed size is selected based on the center of gravity [29]. Hence, the sufficiency of valid fingerprint information contained in the segmented patch is well guaranteed. Besides, four patches around the center of gravity are also selected (see Fig. 1) to enlarge the training sets, preventing network from overfitting. However, these patches require an additional check to filter out those with little information contained. Only when the ratio of fingerprint foreground region is greater than 60%, the local patch will be selected as one training sample. In the testing stage, a score fusion, using average-rule, of the five local patches are performed, which has been proved to benefit the detection accuracy [30]. It is worth noting that by adopting the above patch strategy, the performance and robustness of our CNN-based network against various sizes of input images are improved while only little computational cost is added.

The entire process of preprocessing is presented in Fig. 1. ROI is firstly extracted to eliminate the interference of background areas, and then five local patches with size $112 \times 112$ are selected and segmented as training samples.

B. FLDNet ARCHITECTURES

The existing networks for fingerprint liveness detection (FLD) [24]–[27] suffer from high computational cost and weak generalization ability. For this reason, FLDNet with specially designed block architecture and an attention pooling layer is proposed, targeting a network more lightweight yet powerful.

1) THE WEAKNESS OF GAP IN FLD

Due to the Global Average Pooling’s (GAP) advantages of reducing the parameters and preventing over-fitting [31], it is employed by many state-of-the-art networks, e.g. ResNets [32], DenseNets [33] and most light networks, like MobileNetV2 [34], ShuffleNetV2 [35]. For face tasks, however, researchers [36], [37] have observed that CNNs with GAP layer are less accurate than those without GAP. Further, FeatherNets [37] replace the GAP with specially designed Streaming Module, and achieve significant performance for face anti-spoofing. Inspired by these works, we carry out experiments to evaluate the weakness of GAP in FLD and get the same conclusion (Table 8). The main insight is that GAP treats every unit of the feature map equally, which is not suitable for FLD.

Fig. 2 shows the brief structure of FLDNet. Each unit of the feature map only depends on a certain region of the input image, which is known as the receptive field [38]. RF-center is the receptive field of the center red unit, while RF-edge corresponds to the edge green one. Although RF-center and RF-edge are of the same size theoretically, they are at the different locations of the input image. In our FLD task, the input of the network is segmented local patches (see Fig. 1). Part of the edge regions of these patches is blank, which is of no use for liveness detection. Besides, the texture information in the edge region is incomplete due to the segmentation (e.g. loss of the ridge continuity). Furthermore, for every convolutional

![FIGURE 1. Flow chart of preprocessing. The original fingerprint image is captured by Green Bit scanner in LivDet 2015. Region of interest (ROI) extraction is firstly applied to obtain the fingerprint foreground region. The red patch is located at the center of gravity. Four green patches around will also be selected if the thresholds are satisfied.](image-url)
Y. Zhang et al.: FLDNet: Light Dense CNN for Fingerprint Liveness Detection

FIGURE 2. The structure of our CNN-based FLDNet. The center red unit with a larger effective receptive field is more important than the edge green one. To differentiate this different importance, the attention mechanism is adopted to modify the original GAP.

FIGURE 3. Attention pooling. The weighted average value is calculated for each feature map in the output of the last convolutional layer and aligned as a one-dimensional output feature vector. The weight matrices come from data-driven learning.

layer in FLDNet, each side of the input is zero-padded to keep the feature map size fixed, resulting in a large area of the RF-edge actually contains no fingerprint information. Hence, the center red unit is more important than the edge green one since the RF-center contains more effective information from the input image. However, the “equally treatment” concept of GAP is contrary to the above analysis, which may explain why GAP is not suitable for FLD. This view is further verified through experiments (Table 8).

2) ATTENTION POOLING

To differentiate the importance of units at different positions of the feature map, the attention mechanism is adopted to modify the original GAP. Instead of computing the average value, the weighted average value for each feature map is calculated and aligned as the output of the proposed attention pooling layer (see Fig. 3). The computation expression for the attention pooling layer is represented as:

\[
V_c = \frac{\sum_{i,j} K_{i,j,c} \times FM_{i,j,c}}{W \times H}
\]  

where \( FM_c \) is the \( c \)th input feature map (\( C \) feature maps in total) of size \( W \times H \) (\( W, H \) denote the width and height respectively), \( K_c \) is the weight matrix of size \( W \times H \) for the \( c \)th feature map, \( V_c \) is the weighted average value of units in the \( c \)th feature map, and \((i, j)\) denotes the spatial position in \( FM_c \). Our modification retains the characteristics of GAP that each feature map only contributes to a single element of the output \( V \) (size \( 1 \times 1 \times C \)). Without many parameters or computational costs added, the attention pooling layer still holds the advantages of reducing dimensions and preventing overfitting.

The weight matrices that reflect the importance of units at different positions of the feature map come from data-driven learning instead of being preset by prior knowledge. We visualize the learned weight matrices to prove our analysis of weakness of GAP (see Fig. 9). Note that the attention pooling can be implemented by using a depthwise convolutional layer.

3) NETWORK ARCHITECTURE DETAIL

FLDNet is mainly built based on DenseNet. For convenience, we use the same concepts as those in [33]. \( F_i(\cdot) \) is a non-linear transformation implemented by the \( i \)th layer and \( x_i \) is the corresponding output. Firstly, we introduce two types of paths that are incorporated in the construction of FLDNet.

a: RESIDUAL PATH

ResNets [32] add shortcut connections to bypass the non-linear transformation with an identity mapping:

\[
x_i = F_i (x_{i-1}) + x_{i-1}
\]  

We denote the element-wise addition (channel by channel) as \textit{eltwise operation}.

b: DENSELY CONNECTED PATH

DenseNets [33] introduce direct connections from any layer to subsequent layers:

\[
x_i = F_i ([x_0, x_1, \ldots, x_{i-1}])
\]  

where \([x_0, x_1, \ldots, x_{i-1}]\) is the concatenation of the outputs of all preceding layers. We denote the channel-wise concatenation as \textit{concat operation}. Basically, DenseNets differ from...
ResNets only in that the outputs of the layers are concatenated instead of summed [33].

The concat operation in the middle of the original dense block is replaced by an eltwise operation followed by a $2 \times 2$ average pooling layer. Meanwhile, the layer (Conv $3 \times 3$ - BN-ReLU) before the eltwise operation maintains the number of feature maps, while other layers in the block all produce $k$ (growth rate) feature maps. By doing so, the residual path is integrated into the middle of the original dense block. Each modified block adds $(n - 1)k$ feature maps to the global state, where $n$ is the number of layers contained in the block. We refer to this new block as Block D&R (see Fig. 4). It has been proved in Dual Path Networks [39] that ResNets enable feature reusage but are poor at exploring new features, while DenseNets keep exploring new features but suffer from higher redundancy. Our proposed block structure enjoys the benefits from two complementary network topologies by integrating the residual path and the densely connected path. Besides, transition layers between two Block D&Rs facilitate down-sampling, which consist of a $1 \times 1$ convolutional layer and a batch normalization (BN) layer followed by a $2 \times 2$ average pooling layer. The $1 \times 1$ convolution in each transition layer adds $k$ feature maps for compensation. Additionally, considering the size of the input local patches is relatively small, the stride of the first convolutional layer in FLDNet is set to 1 to achieve a good trade-off between the computational cost and retaining much information as possible. The setting of stride = 2 leads to a poor detection accuracy. This observation was also made in our preliminary study [27].

The detailed FLDNet architectures are shown in Table 1, where the number of layers in a block and the growth rate is set to 5 and 12 respectively. For all convolutions (except those with kernel size $1 \times 1$), each side of the inputs is zero-padded to keep the feature map size fixed. The size of the input image is $112 \times 112$. Before entering the first Block D&R, a convolution with kernel size $5 \times 5$, stride = 1 and 16 output feature maps are performed, then 3 Block D&Rs and 2 transition layers are attached. Each transition layer adds 12 feature maps for compensation. Finally, the attention pooling layer is used without adding a fully connected layer, directly converting the $4 \times 4 \times 184$ feature maps into a one-dimensional vector. After that, a SoftMax classifier is attached for prediction and cross-entropy is used as the loss function. Our primary FLDNet only has 0.48M parameters.

### III. EXPERIMENTS

In this section, preliminary work will be introduced firstly, which includes evaluation metrics, datasets used for training and testing, the data augmentation method and the training strategy. Then, the performance of the proposed method is evaluated by comparative experiments under three different scenarios. Finally, the effectiveness of designed network architecture is further verified by ablation experiments.

#### A. PRELIMINARY WORK

1) **EVALUATION METRICS**

The performance evaluations follow the metrics used in LivDet [40]. Average Classification Error (ACE) is the average of the rate of misclassified live fingerprints ($F_{errlive}$) and the rate of misclassified fake fingerprints ($F_{errfake}$). The ACE applied for the performance evaluation is defined as:

$$ACE = \frac{F_{errfake} + F_{errlive}}{2}$$  \hspace{1cm} (4)
The threshold for fingerprint liveness detection in this paper is set to 0.5. Fingerprint images with a liveness score higher than 0.5 are distinguished as living samples, while those under 0.5 are considered as spoofs. Unless otherwise specified, the following experiments are performed based on this threshold.

2) DATASETS
The performance of the proposed method is evaluated on two public datasets, LivDet 2013 [41] and LivDet 2015 [6]. Considering that the testing sets of LivDet 2017 and 2019 have not been public yet, they are not involved in the evaluation. LivDet 2013 consists of four datasets, Biometrika, ItalData, Swipe and CrossMatch, each captured by the corresponding fingerprint reader. Fingerprint images from the Swipe sensor are considerably different from other datasets since fingers are required to swipe from top to bottom during the collection. This dataset is excluded from our experiments. Besides, the CrossMatch dataset is also removed because anomalies have been found in the fingerprint data from this reader [40]. The fingerprint replicas in LivDet 2013 are made by gelatin, latex, ecoflex, modasil and wood glue using the non-cooperative method. For LivDet 2015, four datasets are included, which are Biometrika, Digital Persona, Green Bit and CrossMatch respectively. Note that new materials are applied to fabricate fake fingerprints in the testing sets of LivDet 2015. More specifically, liquid ecoflex and RTV for Biometrika, Digital Persona, and Green Bit sensors, and gelatin and OOMOO for CrossMatch sensor. Different from LivDet 2013, spoofs in LivDet 2015 are fabricated using the cooperative method.

In LivDet datasets, the sizes of captured images vary on different fingerprint readers, ranging from $252 \times 324$ to $1000 \times 1000$ pixels. Hence, directly sending the whole fingerprint images will require the network structure to change accordingly, while feeding network with fixed-size local patches can help it adapt to this diversity. The detailed information of datasets used in the experiments is summarized in Table 2. Sample images of LivDet 2015 and LiveDet 2013 are shown in Fig. 6 and Fig. 7 respectively.

B. IMPLEMENTATION DETAIL
1) DATA AUGMENTATION
Each local patch is flipped horizontally and vertically, and rotated at four different angles for data augmentation. Except for these standard methods, mixup [42] is also employed to
TABLE 2. Summary of the liveness detection (LivDet) datasets used in the experiments.

| Dataset  | Sensor       | Image Size | Resolution (dpi) | # of training (Live/SpooF) | # of testing (Live/SpooF) | Spoof Materials                  |
|----------|--------------|------------|-----------------|-----------------------------|---------------------------|--------------------------------|
| LivDet 2013* | Biometrika   | 315 × 372  | 569             | 1000/1000                   | 1000/1000                  | Ecoflex, Gelatine, Latex, Modasil, Wood Glue |
|          | Itldata      | 640 × 480  | 500             | 1000/1000                   | 1000/1000                  |                                           |
| LivDet 2015 | Biometrika   | 1000 × 1000| 1000            | 1000/1000                   | 1000/1500                  | Ecoflex, Gelatine, Latex, Wood Glue, Liquid Ecoflex, RTV |
|          | Digital Persona | 252 × 324  | 500             | 1000/1000                   | 1000/1500                  |                                           |
|          | Green Bit    | 500 × 500  | 500             | 1000/1000                   | 1000/1500                  |                                           |
|          | CrossMatch   | 640 × 480  | 500             | 1510/1473                   | 1500/1448                  | Body Double, Ecoflex, PlayDoh, OOMOO, Gelatin |

The training sets do not include the spoof materials in bold.

* The spoofs in LivDet 2013 are fabricated using the non-cooperative method, while those in LivDet 2015 are fabricated using the cooperative method.

TABLE 3. ACE comparisons between FLDNets with different structures on LivDet 2015 dataset (%).

| State-of-the-Art [19] | Growth rate | Number of layers | Biometrika | CrossMatch | Digital Persona | GreenBit | Average |
|-----------------------|-------------|------------------|------------|------------|-----------------|----------|---------|
| FLDNet 1              | 12          | 3                | 3.2        | 2.28       | 4.6             | 1.2      | 2.82    |
| FLDNet 2              | 24          | 3                | 4.31       | 6.56       | 4.28            | 1.95     | 4.28    |
| FLDNet 3              | 12          | 5                | 2.69       | 3.94       | 6.16            | 3.87     | 4.17    |
| FLDNet 4              | 24          | 5                | 2.34       | 1.54       | 2.58            | 0.56     | 1.76    |
| FLDNet 5              | 12          | 7                | 2.99       | 2.26       | 2.46            | 1.15     | 2.22    |

The best performance on each dataset (each column) is bold.

improve the generalization ability of FLDNet. Essentially, mixup is a data-agnostic augmentation method which uses convex combinations of pairs of training samples and their labels to train networks:

\[
\tilde{x} = \lambda x_i + (1 - \lambda) x_j \\
\tilde{y} = \lambda y_i + (1 - \lambda) y_j
\]

where \((\tilde{x}, \tilde{y})\) is the produced virtual samples, \((x_i, y_i)\) and \((x_j, y_j)\) are randomly selected feature-target vectors from the training data \(x\) is the normalized gray image with size \(112 \times 112\) and \(y\) is the corresponding label), and \(\lambda \in [0, 1]\) controls the strength of this linear interpolation.

Mixup is applied within a mini-batch. More specifically, virtual samples are generated online by applying (5) to randomly selected two samples. Besides, applying mixup only in the previous phase of the training performs better than using it throughout the entire process. In our experiments, mixup is applied in the first 25% training iterations. The effectiveness of mixup is further verified by experiments (Table 4).

2) TRAINING STRATEGY

Caffe is used to implement the proposed networks. It initializes the model parameters with Gaussian distributions. Models are trained using Stochastic Gradient Descent (SGD) with batch size 32 for 250000 iterations. The learning rate is initialized at 0.01 and reduced by 20% per 50000 iterations in the training stage. A dropout layer is added after each convolutional layer (except the first one) and the dropout rate is set to 0.2, as shown in Fig. 5.

C. RESULT ANALYSIS

1) COMPARISONS BETWEEN DIFFERENT FLDNET STRUCTURES

To determine the best settings of the growth rate \(k\) and the number of layers contained in a Block D&R, we compare the performance of FLDNets with different structures on LivDet 2015. It is recommended to set the number of layers contained in a Block D&R to an odd number since the residual path is integrated into the middle, while others are original densely connected paths. The comparison results presented in Table 3 show that a relatively small growth rate \((k = 12)\) is sufficient to obtain state-of-the-art results on the datasets we test on. For the number of layers contained in a block, FLDNet 3 with 5 layers in a block achieves the best performance with an average ACE of 1.76% over all sensors. Increasing the depth of the network within a certain range can improve the performance, the overall classification error is significantly reduced comparing FLDNet 3 with FLDNet 1. However, a deeper network structure does not necessarily lead to a better performance for FLDNet according to the testing results of FLDNet 3 and FLDNet 5. the best settings of the growth rate and the number of layers in a block are 12 and 5 respectively, which indicate that FLDNet can achieve a compelling performance with a limited parameter size (0.48M). The fldnets tested in the following experiments all adopt this setting for structure.
TABLE 4. ACE comparisons between FLDNet and existing methods on LivDet 2013 and LivDet 2015 datasets (%).

| Method       | Biometrika | Italdata | Average | Biometrika | CrossMatch | Digital Persona | GreenBit | Average | Overall Average |
|--------------|------------|----------|---------|------------|------------|----------------|----------|---------|----------------|
| WLBD [18]    | 0.4        | 0.95     | 0.68    | 9.64       | 10.82      | 13.72         | 4.53     | 9.68    | 6.68           |
| FPAD [19]    | 0.5        | 0.1      | 0.3     | 3.2        | 2.28       | 4.6           | 1.2      | 2.82    | 1.98           |
| DRN [21]     | 2          | 1.8      | 1.9     | 6.24       | 3.46       | 6.8           | 4.77     | 5.32    | 4.18           |
| VGG-19 [22]  | 1.8        | 0.4      | 1.1     | 5.6        | 1.9        | 6.3           | 4.6      | 4.6     | 3.43           |
| fpADNet [26] | 0.9        | 1.3      | 1.1     | 4.1        | 0.3        | 8.5           | 1.4      | 3.58    | 2.75           |
| Slim-ResCNN [27] | 0.47 | 3.01   | 1.74   | 3.10       | 4.32       | 2.37          | 2.64     | 3.11    | 2.65           |
| ResNet (reported by [27]) | 4.09 | 17.16 | 10.63 | 32.06 | 9.59 | 40.61 | 14.74 | 24.25 | 19.71 |
| DenseNet [33]† | 0.55 | 3.00   | 1.78   | 4.39       | 3.68       | 10.90         | 6.67     | 6.41    | 4.87           |
| FLDNet       | 0.36       | 1.35     | 0.86    | 2.95       | 1.78       | 3.61          | 0.53     | 2.22    | 1.76           |
| FLDNet (mixup)† | 0.25 | 1.15   | 0.70   | 2.34       | 1.54       | 2.58          | 0.56     | 1.76    | 1.40           |
| FLDNet (three)† | 0.30 | 0.20  | 0.25   | 3.72       | 1.41       | 2.17          | 0.48     | 1.94    | 1.38           |

The best performance on each dataset (each column) is bold.
* All the results of DenseNet are run by ourselves. The detailed settings are introduced in ablation experiments.
† Except for the standard data augmentation methods, mixup is also applied in training.
‡ The threshold for classification on the testing set is learned over the training set at the EER point of the ROC curve. Note that mixup is also applied.

2) COMPARISONS WITH STATE-OF-THE-ART METHODS
The proposed method is tested under following three scenarios of FLD, which evaluate an algorithm’s robustness against unknown spoof materials and the use of different sensors.

a: KNOWN-SENSOR AND KNOWN-MATERIAL SCENARIO
In this scenario, same sensor is used to capture all the images for training and testing, and all the materials used to fabricate spoofs in the testing set are known. The ACEs of FLDNet are compared to several existing works, including non-deep learning-based methods [18], [19], deep learning based methods [21], [22], [26], [27] and two baseline models, ResNet [32] and DenseNet [33]. The results presented in Table 4 reveal that FLDNet outperforms the state-of-the-art methods, showing an improvement of 1.06% in the average ACEs over LiveDet 2015 datasets. It is worth noting that the FLDNet has a relatively small parameter size (0.48M), which enables it to run on low specification systems. According to [40], most of the algorithms submitted to LivDet 2015 performed poorly on Digital Persona sensor because of the small image size. Our patch strategy reduces the impact of this limitation to some extent. Besides, mixup is proved to boost the detection accuracy with an average improvement of 0.36%. It is applied in all the following experiments.
Further, the optimal threshold is learned over the training set at the equal error rate (EER) point of the ROC (Receiver Operating Characteristics) curve. Although the effectiveness of optimal threshold differs from sensors, the increment of the overall performance is not obvious (Table 4). The ROC curves of FLDNet on LivDet 2015 are shown in Fig. 8. The false positive rate and the true positive rate are zoomed to make the trend more clearly.

b: KNOWN-SENSOR AND UNKNOWN-MATERIAL SCENARIO
In this setting, fingerprint images in the training and testing sets are captured by the same sensor. But new materials that are unseen during training are used to fabricate spoofs in the testing set. Since the testing sets of LivDet 2015 contain fake fingerprints made of unknown materials (Table 2), we firstly verify the robustness of the proposed method in cross-material scenario on LiveDet 2015 dataset. A detailed performance comparison between FLDNet and the LivDet 2015 winner [6] and the LivDet 2017 winner [27], is shown in Table 5. Ferrfakeknown and Ferrfakeunknown are the percentages of misclassified spoofs fabricated by known and previously unseen materials respectively. Ferrfake is an average of Ferrfakeknown and Ferrfakeunknown, weighted by the number of samples in each category. Ferrlive is the percentage of misclassified live fingerprints. The comparison results present that FLDNet outperforms two winners in nearly all the metrics (except for Ferrfakeunknown on CrossMatch) on four different datasets. Especially the improvement of Ferrfakeunknown (2.14% on average), proves that FLDNet

FIGURE 8. ROC curves of FLDNet on LiveDet2015 datasets (Biometrika, Digital Persona, Green Bit and CrossMatch). The range of false positive rate and true positive rate are adjusted to make the trend more clearly.
is robust facing cross-material scenario. To further verify the robustness of FLDNet against unknown spoof materials, another set of experiments are performed on LivDet 2013 datasets, following the protocol adopted by [22]. The results presented in Table 6 show that FLDNet achieves an average ACE of 0.91\%, which is comparable to state-of-the-art accuracy.

### ABLATION EXPERIMENTS

Several ablations are executed to verify the effectiveness of the proposed network architecture, shown in Table 8.

#### WHY MODIFY GAP LAYER WITH THE ATTENTION MECHANISM

In Section II(B), we analyze the weakness of GAP by pointing out the “equally treatment” concept is not suitable for FLD. This view is verified through comparative experiments (Table 8) on two network architectures. DenseNet AP, where GAP layer is replaced by the attention pooling layer, outperforms the original one with an improvement of 2.99\% in average ACE. Besides, by comparing FLDNet A with

| Method                        | LiveDet2015 | FERLinve (%) | FERLfake (%) | FERLfake known* (%) | FERLfake unknown* (%) | ACE (%) |
|-------------------------------|-------------|--------------|--------------|---------------------|-----------------------|--------|
| LivDet 2015 winner [6]        | Biometrica  | 8.50         | 3.73         | 2.70                | 5.80                  | 5.64   |
|                               | CrossMatch  | 0.93         | 2.90         | 2.12                | 4.02                  | 1.90   |
|                               | Digital Persona | 8.10     | 5.07         | 4.60                | 6.00                  | 6.28   |
|                               | GreenBit    | 3.50         | 5.33         | 4.30                | 7.40                  | 4.60   |
| Average                       |             | 4.78         | 4.27         | 3.48                | 5.72                  | 4.49   |
| LivDet 2017 winner [27]       | Biometrica  | 3.55         | 4.23         | 2.44                | 7.72                  | 2.98   |
|                               | CrossMatch  | 1.72         | 3.91         | 3.18                | 4.58                  | 2.99   |
|                               | Digital Persona | 4.28     | 4.78         | 3.73                | 6.40                  | 4.58   |
|                               | GreenBit    | 2.22         | 2.65         | 2.42                | 3.11                  | 2.19   |
| Average                       |             | 3.14         | 3.89         | 2.99                | 5.45                  | 3.18   |
| Proposed Method               | Biometrica  | 1.01         | 3.67         | 2.61                | 5.75                  | 2.34   |
|                               | CrossMatch  | 0.87         | 2.21         | 0.69                | 4.45                  | 1.54   |
|                               | Digital Persona | 3.26     | 1.90         | 1.43                | 2.85                  | 2.58   |
|                               | GreenBit    | 0.71         | 0.41         | 0.52                | 0.21                  | 0.56   |
| Average                       |             | 1.46         | 2.05         | 1.31                | 3.31                  | 1.76   |

\* FLDNet outperforms two winners except for FERLfakeunknown on CrossMatch (increases 0.43\% compared to the LivDet 2015 winner).

\*\* FERLfakeunknown is the percentage of misclassified spoof images fabricated by known materials.

\*\* FERLfakeunknown is the percentage of misclassified spoof images fabricated by new materials that are previously unseen during the training.

### TABLE 6. ACE comparisons between FLDNet and existing methods on LivDet 2013 dataset for cross-material experiments (%).

| Dataset           | Materials (training) | Materials (testing) | VGG-19 [22] | TripletNet [24] | FSB [25] | FLDNet |
|-------------------|----------------------|---------------------|-------------|-----------------|---------|-------|
| Biometrica        | Modasil, Wood Glue   | Ecoflex, Gelatine, Latex | 4.9         | 2.1             | 1.3     | 0.87  |
| Italdata          |                      |                     | 6.3         | 1.25            | 0.6     | 0.94  |
| Average           |                      |                     | 5.6         | 1.68            | 0.95    | 0.91  |

The best performance on each dataset (each row) is **bold**.

### TABLE 7. ACE comparisons between FLDNet and existing methods on LivDet 2013 dataset for cross-sensor experiments (%).

| Training Dataset | VGG-19 [22] | TripletNet [24] | FSB [25] | FLDNet |
|------------------|-------------|-----------------|---------|-------|
| Biometrica       | 8.8         | 1.55            | 4.30    | 2.10  |
| Italdata (Biometrica) | 2.3     | 3.8             | 3.50    | 2.90  |
| Average          | 5.55        | 2.68            | 3.90    | 2.50  |

The best performance on each dataset (each row) is **bold**.
TABLE 8. ACE comparisons on LivDet 2015 dataset for ablation experiments.

| Model          | GAP | Residual path | FC | Biometrika | CrossMatch | Digital Persona | GreenBit | Average |
|----------------|-----|---------------|----|------------|------------|-----------------|----------|---------|
| DenseNet [33]  | √   | ×             | ×  | 4.39       | 3.68       | 10.90           | 6.67     | 6.41    |
| DenseNet (AP)† | ×   | ×             | ×  | 3.78       | 2.77       | 5.40            | 1.72     | 3.42    |
| FLDNet A       | √   | √             | ×  | 2.90       | 5.06       | 4.56            | 0.90     | 3.36    |
| FLDNet B       | ×   | √             | √  | 2.34       | 1.54       | 2.58            | 0.56     | 1.76    |
| FLDNet C       | ×   | √             | √  | 2.81       | 5.37       | 4.03            | 0.78     | 3.25    |

FC and GAP represent applying Fully Connected layer and Global Average Pooling layer to replace the proposed attention pooling layer respectively. Residual path means using the proposed Block D&R to replace the original dense block.

* All the results of DenseNet are run by ourselves, where the stride of the first convolution is set to 1.

† The GAP layer of the original DenseNet is replaced by the proposed Attention Pooling layer (AP).

FIGURE 9. Heat-maps of average (absolute) value of weight matrices (size 4 × 4) in the trained FLDNet on four datasets in LivDet 2015, which are Biometrika (a), Digital Persona (b), GreenBit (c) and CrossMatch (d) respectively. A square in lighter color indicates that the unit at that position of the feature map receives more attention.

FLDNet B and C, it shows that GAP layer performs worse than the attention pooling layer and fully connected layer. We further investigate if a trained FLDNet does take advantages of the attention mechanism. For each dataset in LivDet 2015, we compute the average (absolute) value of weight matrices (size 4 × 4) in the trained FLDNet respectively. A square in lighter color indicates the unit at that position of the feature map receives more attention. Since the weights come from data-driven learning, a unit receiving more attention suggests that it is more valuable for liveness detection. The central areas of heatmaps are generally lighter in color than the outer areas (see Fig. 9), which proves our previous analysis that the receptive fields of the central units contain more effective information.

b: WHY ADD RESIDUAL PATH TO DENSE BLOCK
In the design of Block D&R, residual path is integrated into the original dense block to further improve the performance by benefitting from two complementary network structures. The effectiveness of this design is proved by comparing the performance of DenseNet with FLDNet (Table 8). To ensure a fair comparison, we eliminate all other factors such as differences in optimizer settings and data preprocessing by adopting the publicly available Caffe implementation for DenseNet by [33]. Especially, the stride of the first convolution is modified to 1, since the required input image size of the original DenseNet is 224 × 224, while the size of local patches utilized in this experiment is 112 × 112. Besides, this setting is necessary to ensure a fair comparison since we have mentioned in Section II that stride = 2 for the first convolution will lead to a poor performance for fingerprint anti-spoofing tasks. Comparing DenseNet with FLDNet A or DenseNet AP with FLDNet B (Table 8), adding residual path to the original dense block can effectively improve the overall performance.

c: WHY NOT USE FULLY CONNECTED LAYER
Fully Connected layer (FC) is an available choice for replacing GAP, since it also treats units at different positions differently. The results in Table 8 suggest that FC does perform better than GAP, however, it doesn’t reduce the overall ACE significantly compared to the proposed attention pooling layer (comparing FLDNet C with FLDNet A and B respectively). Besides, a FC layer is computationally expensive and easily overfitted.

4) PROCESSING TIMES
The processing times of the proposed method are evaluated on a single Nvidia GeForce GTX 1080 GPU. FLDNet takes 7.5ms to process a single local patch on average, which is nearly three times faster than the LivDet 2017 winner as it requires 20ms to process a same size (112 × 112) patch in the same testing environment as reported in [27]. For a single input image, the average classification time, including local patch segmentation, calculating the liveness scores for local patches and producing the final decision, is approximately 18.6ms. Compared to FSB [25], which requires 100ms on a single Nvidia GeForce GTX 1080 Ti GPU, FLDNet can better satisfy the real-time processing requirements.
IV. CONCLUSION
We propose a lightweight network architecture (FLDNet) with attention pooling layer, which achieves state-of-the-art accuracy for fingerprint liveness detection at a relatively low computational complexity. We point out the limitation of GAP in fingerprint anti-spoofing tasks and overcome it by applying the attention mechanism. A new block structure (Block D&R) is put forward, where the residual path is integrated into the original dense block. The comparative experiments prove that FLDNet achieves state-of-the-art accuracy on public LivDet datasets, and is more robust in cross-material- and cross-scenario sensors. Additionally, FLDNet, with 0.48M parameters, can be applied on low specification systems due to its compactness. The directions of our future research are further improving the network’s performance on small size fingerprints and robustness against unknown spoof materials.

REFERENCES
[1] N. Evans, ‘Handbook of Biometric Anti-spoofing: Presentation Attack Detection’. Cham, Switzerland: Springer, 2019. [Online]. Available: https://www.springer.com/gp/book/9783319926261
[2] S. S. Arora, K. Cao, A. K. Jain, and N. G. Paulter, “Design and fabrication of 3D fingerprint targets,” IEEE Trans. Inf. Forensics Security, vol. 11, no. 10, pp. 2284–2297, Oct. 2016.
[3] P. V. Reddy, A. Kumar, S. M. K. Rahman, and T. S. Mundra, “A new anti-spoofing approach for biometric devices,” IEEE Trans. Biomed. Circuits Syst., vol. 2, no. 4, pp. 328–337, Dec. 2008.
[4] D. Baldisserra, A. Franco, D. Maio, and D. Maltoni, “Fake fingerprint detection by odor analysis,” in Proc. Int. Conf. Biometrics. Berlin, Germany: Springer, Jan. 2006, pp. 265–272. [Online]. Available: https://link.springer.com/chapter/10.1007/11608288_36
[5] S. A. Abhayankar and S. C. Schuckers, “A wavelet-based approach to detecting liveness in fingerprint scanners,” in Proc. Int. Soc. Opt. Photon., Biometric Technol. Hum. Identificat., vol. 5404, Aug. 2004, pp. 278–286.
[6] G. L. Marcialis, F. Roli, and A. Tidu, “Analysis of fingerprint pores for vitality detection,” in Proc. 20th Int. Conf. Pattern Recognition (ICPR), Istanbul, Turkey, Aug. 2010, pp. 23–26.
[7] E. Marasco and C. Sansone, “Combining perspiration- and morphology-based static features for fingerprint liveness detection,” Pattern Recognit. Lett., vol. 33, no. 9, pp. 1148–1156, Jul. 2012.
[8] A. Antonelli, R. Cappelli, D. Maio, and D. Maltoni, “Fake fingerprint detection by skin distortion analysis,” IEEE Trans. Inf. Forensics Security, vol. 1, no. 3, pp. 360–373, Sep. 2006.
[9] G. Galbally, S. Macrì, and J. Fierrez, “Image quality assessment for fake biometric detection: Application to iris, fingerprint, and face recognition,” IEEE Trans. Image Process., vol. 23, no. 2, pp. 710–724, Feb. 2014.
[10] Z. Xia, R. Lv, and X. Sun, “Rotation-invariant Weber pattern and Gabor feature for fingerprint liveness detection,” Multimedia Tools Appl., vol. 77, no. 14, pp. 18187–18200, Jul. 2018.
[11] S. B. NIKAM and S. AGARWAL, “Texture and wavelet-based spoof fingerprint detection for fingerprint biometric systems,” in Proc. 1st Int. Conf. Emerg. Trends Eng. Technol., 2008, pp. 675–680.
[38] W. Luo, Y. Li, R. Urtasun, and R. S. Zemel, “Understanding the effective receptive field in deep convolutional neural networks,” in Proc. Adv. Neural Inf. Process. Syst., 2016, pp. 4898–4906.

[39] Y. Chen, J. Li, H. Xiao, X. Jin, S. Yan, and J. Feng, “Dual path networks,” in Proc. NIPS, 2017, pp. 4467–4475.

[40] L. Ghiasi, D. A. Yambay, V. Mura, G. L. Marcialis, F. Roli, and S. A. Schuckers, “Review of the fingerprint liveness detection (LivDet) competition series: 2009 to 2015,” Image Vis. Comput., vol. 58, pp. 110–128, Feb. 2017.

[41] D. Yambay, J. S. Doyle, K. W. Bowyer, A. Czajka, and S. Schuckers, “LivDet-iris 2013–Iris liveness detection competition 2013,” in Proc. IEEE Int. Joint Conf. Biometrics, Madrid, Spain, Sep. 2014, pp. 1–6.

[42] H. Zhang, Y. N. Dauphin, and D. LopezPaz, “Mixup: Beyond empirical risk minimization.” 2017, arXiv:1710.09412. [Online]. Available: https://arxiv.org/abs/1710.09412

YONGLIANG ZHANG received the B.S. and M.S. degrees from Jilin University, China, in 2000 and 2003, respectively, and the Ph.D. degree from Shanghai Jiao Tong University, China, in 2007. He is currently an Associate Professor with the College of Computer Science and Technology, Zhejiang University of Technology. His research interests include biometric identification, pattern recognition, and artificial intelligence.

SHENGYI PAN was born in Hangzhou, China, in 1999. He is currently pursuing the B.S. degree in computer science with Zhejiang University of Technology, China. His research interests include biometric identification, pattern recognition, and machine learning.

XIAOSI ZHAN received the B.S. and M.S. degrees from Jilin University, China, in 1998 and 2001, respectively, and the Ph.D. degree from Nanjing University, China, in 2004. He is currently a Professor with the School of Science and Technology, Zhejiang International Studies University. His research interests include image processing, pattern recognition, and biometric identification.

ZHIIWEI LI received the M.S. degree from the Center for Optics and Optoelectronics Research, College of Science, Zhejiang University of Technology, China, in 2016. He is currently an Algorithm Engineer with Hangzhou Jing Lianwen Technology Company Ltd. His research interests include biometric identification, machine learning, and digital forensic.

MINGHUA GAO received the B.S. degree from the Tongda College, Nanjing University of Posts and Telecommunications, China, in 2018. He is currently pursuing the M.S. degree in computer science with the Zhejiang University of Technology. His research interests include image processing and machine learning.

CHENHAO GAO received the B.S. degree from the College of Computer Science and Technology, Zhejiang University of Technology, China, in 2018, where he is currently pursuing the M.S. degree in software engineering. His research interests include image processing, biometric identification, and machine learning.

* * *