EmotioNet: A 3-D Convolutional Neural Network for EEG-based Emotion Recognition

Yi Wang, Zhiyi Huang, Brendan McCane
Department of Computer Science
University of Otago
Dunedin, New Zealand
{davidwang, hzy, mccane}@cs.otago.ac.nz

Phoebe Neo
Department of Psychology
University of Otago
Dunedin, New Zealand
phoebe@psy.otago.ac.nz

***Abstract*—In this paper, an emotional EEG-specific three-dimensional Convolutional Neural Network, EmotioNet, is proposed and implemented to accurately recognize emotion states. For the first time, raw data in the benchmark emotional EEG database, i.e. DEAP, are used as the input to a CNN architecture. In order to investigate the spatio-temporal character of emotional features, the effectiveness of 2-D and 3-D convolution kernels, which extract spatial and temporal features separately and simultaneously, are compared in detail. Furthermore, two major problems of EEG-based emotion recognition, namely, covariance shift and the unreliability of emotional ground truth, are described, and the effectiveness of batch normalization and dense prediction, which alleviate these problems respectively, are also investigated. Experimental results show that 3-D kernels, batch normalization, and dense prediction are all essential techniques for the emotional EEG-specific CNN architecture. The proposed EmotioNet, namely, a 3-D covariance shift adaptation-based CNN with a dense prediction layer, achieves classification rates of 73.3% and 72.1% for arousal and valence, equivalent to the best performance of several previous studies. Importantly, our results are based on automatic feature extraction, which is in contrast to previous handcrafted features. Therefore, EmotioNet provides a new method for EEG-based emotion recognition.**

***Index Terms*—Emotion recognition, EEG, 3-D CNN, spatio-temporal emotional features, covariance shift, the unreliability of emotional ground truth**

I. Introduction

As a critical emotional modality, the electroencephalogram (EEG) is capable of representing the inner emotional states without the influence of subjective human control. Specifically, an emotion-related process will involve a change in brain activity which can be assessed by non-invasive low-cost scalp EEG recordings [1] [2], thus providing a reasonable way to recognize emotions.

DEAP, which is the benchmark dataset for EEG-based emotion recognition, provides a standardized way to compare the performance of different emotion recognition methods [3]. Several studies have been conducted on EEG-based emotion recognition using DEAP. Conventionally, time or frequency domain features in EEG were extracted manually. Meanwhile, classifiers such as Support Vector Machine (SVM) and K-Nearest Neighbors (KNN) were implemented [4] [5] [6]. Based on these studies, we propose an interesting and challenging question: instead of using handcrafted emotional features, can automatic feature extraction improve recognition performance?

Theoretically, CNN could provide a promising performance due to its advantages of automatic feature detection and hierarchical feature extraction. Therefore, in this paper, we propose EmotioNet, an emotional EEG-specific CNN architecture. Specifically, the network employs raw EEG signals to automatically extract high-level features, and then utilizes these features to predict emotion states. In order to obtain an objective evaluation, the performance of EmotioNet is compared with previous studies on DEAP. Generally, instead of using handcrafted features, we expect EmotioNet to provide a novel method by using automatic emotional feature detection.

To design the architecture of EmotioNet, there are three critical problems that need to be solved, namely: 1) spatio-temporal emotional feature extraction; 2) the covariance shift problem in EEG scenarios; and, 3) the unreliability of emotional ground truth.

Spatio-temporal emotional feature extraction. Only a handful of studies have focused on automatic EEG feature extraction using CNN. In particular, two generic architectures, i.e. EEGNet [7] and Braindecode [8], were designed to extract and classify features in Event Related potential (ERP) and Motor Imagery (MI) EEG signals. Critically, an alternative feature extraction scheme that extracts spatial and temporal features separately was implemented in both architectures, which provided a reference for our emotional EEG feature extraction. However, some studies also reported a simultaneous emotional pattern variation in spatial and temporal domains [9] [10], which presented another reference for the automatic emotional feature extraction. Therefore, our question is: in terms of EEG-based emotional feature extraction, should our network extract spatial and temporal features separately or simultaneously?

In this paper, in order to investigate the spatio-temporal character of emotional feature extraction, we implement a two-dimensional Convolution Neural Network (2-D CNN) according to EEGNet and Braindecode, which alternatively extracts spatial and temporal features, and a three-dimensional Convolution Neural Network (3-D CNN), which simultaneously detects spatial and temporal features.

The covariance shift problem in EEG scenarios. Theoretically, covariance shift, which describes different statistical

978-1-5090-6014-6/18/$31.00 ©2018 IEEE

distributions between training and testing datasets, affects the performance of a recognition model [11]. In emotional EEG scenarios, the influence might be even more severe due to the EEG characteristics of non-stationarity and inter-subject variability [12] [13]. To alleviate this problem, batch normalization, which normalizes the distribution of each feature in every layer [14], is embedded in the network and its effectiveness is fully investigated in Section V.

The unreliability of emotional ground truth. In EEG recognition tasks, in order to augment the amount of training data, one EEG trial is often segmented into several input epochs. In emotional EEG scenarios, for a certain trial, due to the intensity fluctuation of the induced emotion [15], different input epochs of this trial should correspond to their ground truth intensity labels. However, practically only the self-reported state, i.e. the average intensity, is recorded and allocated to all the input epochs of the trial. Consequently, the difference between the average and the ground truth intensity causes the unreliability for each epoch, and influences the training of the model.

To alleviate this problem, instead of training the network on every single epoch and its unreliable label, dense prediction is implemented to train a network on several sequential epochs simultaneously with high efficiency [16]. Since these epochs contain enough general emotion information of the trial, the average prediction would represent the average intensity label more accurately. The effectiveness of dense prediction is presented in Section V.

In this paper, we introduce EmotioNet, a 3-D covariance shift adaptation-based CNN architecture with a dense prediction layer, to solve the specific problems in EEG scenarios. We expect this system to provide a novel method for EEG-based emotion recognition. The structure of the paper is as follows: in Section II, a background is presented followed by an introduction of the characteristics of emotional EEG signals in Section III. In Section IV, two types of CNN architectures are employed to recognize different types of emotional EEG patterns, meanwhile, a covariance shift adaptation and dense prediction are also described. The experimental results are given in Section V followed by a discussion and a conclusion in Section VI and VII.

II. Related works

A. DEAP dataset

DEAP is the benchmark dataset for EEG-based emotion recognition. Specifically, for each participant, 40 videos with affective tags were presented in turn to induce specific emotions. During each video, EEG, physiological and video signals were collected, then, a self-reported emotion state using a valence-arousal-dominance space [17] was evaluated immediately after the video. The preprocessing steps were conducted after the data and label collection. Specifically, the data were downsampled to 128Hz, and the electrooculogram (EOG) artifacts were removed. Then the data were filtered in the frequency range of 4.0Hz to 45.0Hz.

In our task, only EEG signals and their corresponding valence-arousal states from 32 participants are taken as data and labels. In detail, *valence* describes the extent of pleasantness with regards to stimuli. It is presented by a continuous value ranging from 1 (negative) through 5 (neutral) to 9 (positive). Meanwhile, *arousal* represents the degree of being awake to stimuli, which has the same range where 1 and 9 indicates passive and active respectively. According to Russell's theory [18], as depicted in Fig. 1, emotions can be expressed by using these two orthogonal states. Therefore, instead of classifying a specific emotion such as excited, valence and arousal are usually taken as basic emotion components that need to be recognized.

B. Emotion recognition using DEAP dataset

Several studies have been conducted on the DEAP dataset, and feature extraction methods have been proposed with different performance. For example, *Chung and Yoon* proposed a Bayes classifier with a weighted-log-posterior function to classify emotion states [19]. Spectral power features in theta (4-8 Hz), alpha (8-12 Hz), beta (12-30 Hz), and gamma (above 30 Hz) bands were used as emotion features. Results showed that the classification rates were 66.6% and 66.4% for valence and arousal respectively. *Rozgic, Vitaladevuni and Prasad* proposed a segment-level feature extraction method with segment-to-response level feature transformation [20]. By using a SVM classifier, accuracies of 76.0% and 68.9% were achieved for valence and arousal. In the study conducted by *Atkinson and Campos*, the minimum-Redundancy-Maximum-Relevance (mRMR) feature selection method was proposed, and a SVM emotion classifier was implemented with the performance of 73.14% (valence) and 73.06% (arousal) [21]. *Gupta and Falk* employed graph theoretical features with a SVM classifier, with a mRMR algorithm to rank the features. The accuracies for valence and arousal were 63% and 66% [22]. In the work presented by *Naser and Saha*, a dual-tree complex wavelet packet transform (DT-CWPT) was adopted as feature extraction method, and a SVM classifier was used

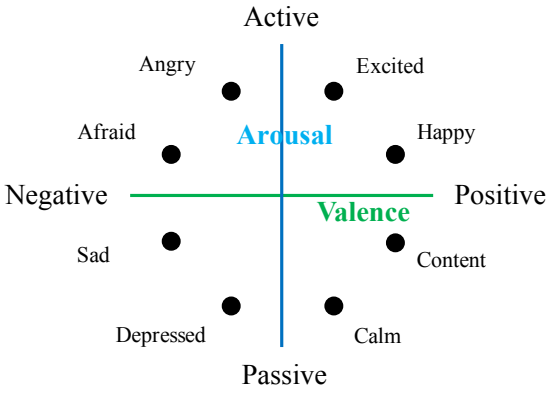

Fig. 1. Russell's valence-arousal emotional model.

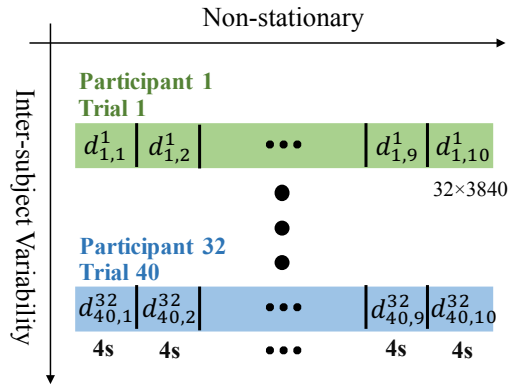

Fig. 2. The segmentation of EEG trials for different participants.

with classification rates of 64.3% and 66.2% [23]. Above all, manually extracted features were mainly adopted in EEG-based emotion recognition in DEAP. However, the effectiveness of automatic feature detection through a Deep Learning architecture, e.g. CNN, still need to be investigated.

C. A generic CNN architecture for EEG pattern recognition

Generally, two successful EEG-specific architectures have been proposed, i.e. EEGNet [7] and Braindecode [8]. In EEGNet, for Layer 2 and 3, the performance of different kernel shapes with the same learning ability (the same number of trainable parameters) was compared in detail. Then, a Spatial-Temporal-Spatial feature extraction architecture, which performed best over four different datasets, i.e. ERP and motor imagery datasets, was suggested by the authors. Meanwhile, Braindecode aimed to detect motor imagery patterns in EEG and employed a Temporal-Spatial-Temporal architecture. According to these studies, this EEG-specific modification in alternate spatio-temporal detection significantly improves the performance of recognition, which provides a reference for our emotion-specific architecture.

III. The Characteristics of Emotional EEG Signals

A. Covariance shift

Covariance shift is a problem when the EEG pattern recognition model is trained on one distribution and then tested on another, which severely affects the performance of the model.

Conventionally, as shown in Fig. 2, in order to augment the amount of data in EEG-specific recognition tasks, all the trials need to be segmented into several input epochs with a given time length, e.g. 4 seconds. The statistical distribution for any epoch can be expressed as follows:

$$\boldsymbol{d}^p_{t,s} \tag{1}$$

where p is the participant number, t is the trial index and s is the epoch index. The ranges for these values are $p \in (1, 32)$, $t \in (1, 40)$, $s \in (1, 10)$. For example, $d^3_{1,5}$ represents the statistical distribution of the 5th epoch in trial 1 for participant 3.

Ideally, all the epochs should have the same distribution, however, due to two major issues, namely, non-stationarity and inter-subject variability, different epochs may have different distributions, which causes the covariance shift problem in EEG recognition scenarios.

1) Non-stationarity: Non-stationarity is an issue when the distribution of an EEG trial changes over time, as described in (2).

$$d^p_{t,s_i} \neq d^p_{t,s_j} \tag{2}$$

where $s_i \neq s_j$. It can be caused by mental changes of a participant, e.g. fatigue, or technical changes, e.g. drying electrode gel. Therefore, for any trial of any participant, the distributions of different epochs might be different. For example, we calculate the epoch distribution in the 3rd trial of participant 18. Each epoch is a matrix with 32 channels and 512 time samples, i.e. a shape of 32×512, and data from all the channels are used to calculate the distribution. As shown in Fig. 3, the distribution of the 1st epoch changes gradually to that of the 8th segment.

2) Inter-subject variability: Inter-subject variability describes the character that for different participants, the statistical distributions are different, as shown in (3).

$$d^{p_i}_{t,s} \neq d^{p_j}_{t,s} \tag{3}$$

where $p_i \neq p_j$, which means the distributions variant within different participants. An example of inter-subject variability is depicted in Fig. 4. We calculate the distribution of a participant using all the data from all the 32 channels. As shown in the figure, the calculated distributions for participant 12 and 10 are significantly different. Assuming they are assigned to training and testing datasets separately, it would severely affect the performance of the model.

B. The unreliability of emotional ground truth

During the recording of a certain emotional EEG trial, a visual stimulus with a specific affective tag was presented to the participant. Ideally, the intensity of the induced emotion

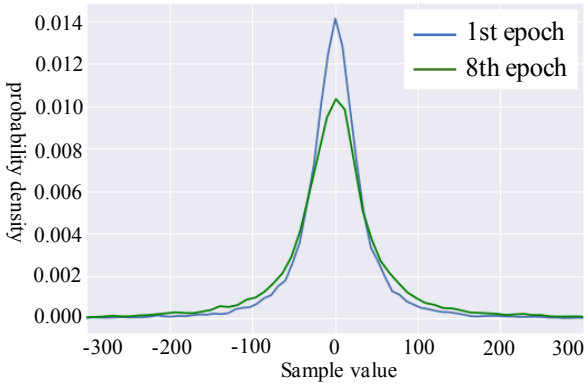

Fig. 3. An example of non-stationarity in DEAP dataset. For the 3rd trial of participant 18, the distributions of the 1st and 8th epochs.

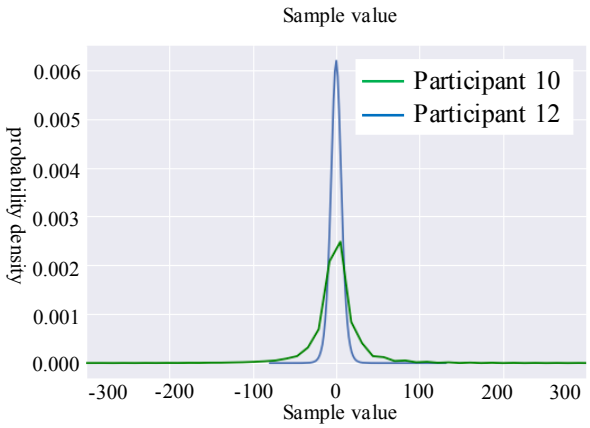

Fig. 4. An example of inter-subject variability in DEAP dataset.

should be constant over time. However, practically due to the subjective sensitivity to the stimulus, the intensity fluctuated over time. A valence intensity fluctuation is shown in Fig. 5.

As illustrated in the figure, in terms of a segmented epoch, the ground truth emotional label should be the averaged intensity over the segmented period. However, conventionally only the averaged intensity over the whole trial, i.e. the self-reported general emotional state, was collected. Therefore, instead of using the ground truth, the general emotional state was allocated to the epoch unit. The intensity difference between the ground truth emotional label and the general state leads to the unreliability of emotional ground truth, which affects the training of the model.

IV. Methods

To solve the problems we address in Section I, 3-D CNN architecture, covariance shift adaptation, and dense prediction are described in this section. The architecture of EmotioNet, which is based on these techniques, is depicted in Fig. 7.

A. 2-D and 3-D CNN architecture

First, a 2-D CNN architecture, which employs a Temporal-Spatial-Temporal feature extraction scheme, is implemented. In detail, the input data has a 2-D shape of channels × time samples. The kernel sizes of (1,10), (32,1), (1,10) are set for Layer 1, 2, and 3 to extract temporal and spatial features alternately. Basically, the 2-D CNN refers to the alternative feature extraction scheme suggested by EEGNet and Braindecode, except that the kernel sizes are different. We take this 2-D architecture as a classic EEG feature extraction scheme, which will be compared with our proposed 3-D architecture in Section V.

In terms of the input dimension, intuitively, the 2-D input format lost the topological position information of the electrodes. To include this information in the input, we reshaped the 2-D matrices (channels × time samples) to 3-D tensors (2-D electrode topological structure × time samples), as illustrated in Fig. 6. As we can see, for each time sample point, the data from each electrode is relocated to its topological position, which provides more information for the model.

Based on this 3-D input format, the 3-D CNN architecture is proposed and implemented, as shown in Fig. 7. Specifically, in Layer 1 and 2, 3-D convolution is utilized to extract spatial and temporal features simultaneously, then, in Layer 3, a spatial fusion, which combines all the spatial characters of these high-level features, is employed. Consequently, the output of this layer only has temporal character. In Layer 4 and 5, temporal feature extraction is integrated in the network to explore high-level temporal features. In Layer 6, an optional dense prediction is embedded, then a Softmax layer is deployed as the output. In each layer, batch normalization is taken as an optional technique whose effectiveness is investigated. Besides, dropout, which randomly drops out network units to avoid overfitting, is employed in Layer 2, 4, and 5.

In terms of kernel sizes, the design idea is that, through all the kernels in different layers, namely, from shallow to deep layers, the coverage for the 2-D electrode topology and the time domain of the input should increase gradually. Specifically, for Layer 1, the kernel size is 2×2×10, which means each spatio-temporal feature output of a kernel is generated by using the input with a local topology of 2×2 and a short time period of 10 sample points. For Layer 2, the kernel size is 2×2×10 with a stride of 2×2×1. However,

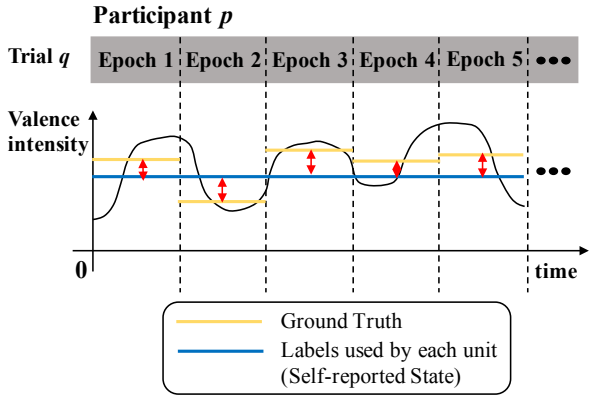

Fig. 5. The unreliability of emotional ground truth.

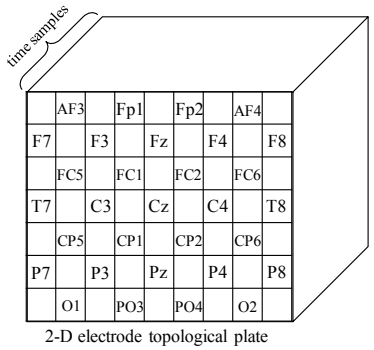

Fig. 6. The 3-D input structure.

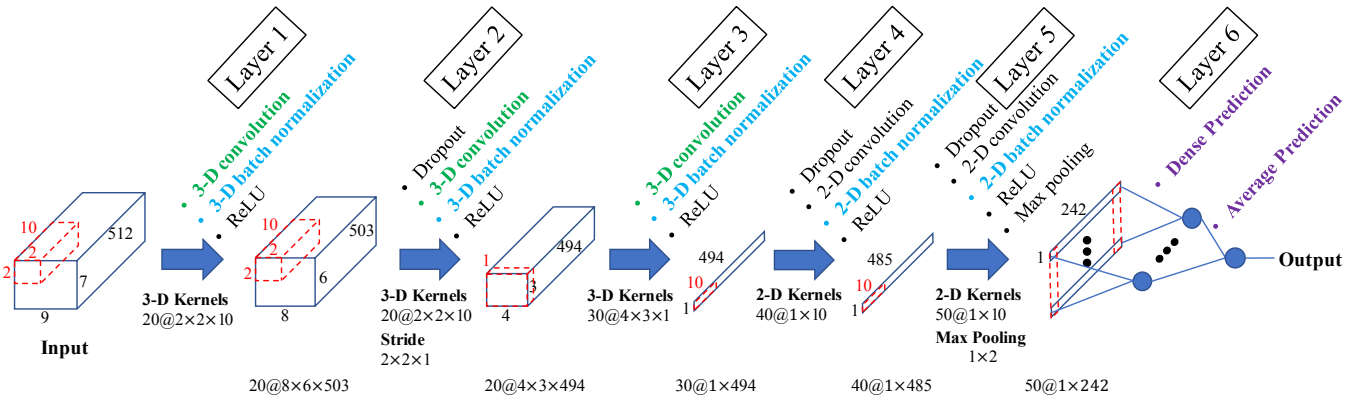

Fig. 7. The architecture of EmotioNet.

this kernel size and stride are based on the output of Layer 1. That means the kernel coverage for the input is broader in both electrode topology and time domain, with a topology coverage of 3×3 and a time domain coverage of 19 sample points. For layer 3 with a kernel shape of 3×4×1, each kernel combines all the spatio-temporal features generated from the previous layer to form a high-level feature. For Layer 4 and 5, a kernel size of 1×10 is used to gradually integrate the time domain information and generate high-level features.

B. Covariance shift adaptation

In this paper, batch normalization is implemented in the network to alleviate the covariance shift problem. Since normalization is an effective data processing method for unifying the feature distributions of the input, a layer-wise trainable normalization, i.e. batch normalization, is also used to normalize features for each layer. In this paper, the effectiveness of batch normalization in emotional EEG recognition scenarios is investigated in Section V.

C. Dense prediction

For emotional EEG recognition tasks, the input epochs with longer segmented window sizes represent the self-reported labels more accurately than the epochs with shorter window sizes. Specifically, the long window sizes cover enough global emotional information of the trials. However, the shorter window sizes only contain local emotional information. Therefore, instead of using each local epoch and its unreliable label to train the network, longer segmented epochs should be taken into consideration. According to the research conducted by *Candra et al*, the window size of four seconds should be a reasonable length that contains enough emotion information [24]. However, for our task, an input epoch with four second time length would lead to a significant increase in trainable parameters, resulting in overfitting and low training efficiency.

To address this problem, dense prediction is implemented to efficiently provide the network with a general emotional information. As shown in Fig. 8, the signal with a time length of four seconds is taken as the input. However, instead of predicting the emotional state using a fully-connected layer directly, dense prediction layer is used to make a time-varying emotion state prediction. Specifically, the receptive field of each output of the dense prediction layer is a window with about two seconds. Each output value of the dense prediction layer, which is calculated on two second segment, can be interpreted as the local prediction. Therefore, the outputs of the dense prediction layer can be taken as a time-varying emotion state. Finally, an average prediction is implemented to form the general emotion state for four second EEG. Comparing with the network using fully-connected layer, the network using dense prediction only has 56562 trainable parameters, which reduces 11858 trainable parameters and efficiently covers the general emotional information of a trial.

V. Results

In this section, we compare the performance of the 2-D and 3-D CNN architectures with different configurations, which are implemented by using PyTorch[1] and are deployed on a GeForce GTX TITAN X GPU. Then, the performance of our proposed EmotioNet is compared with previous work.

First of all, in order to investigate the effectiveness of different techniques, i.e. 3-D CNN architecture, dense prediction, and batch normalization, the architectures with different configurations are tested on the DEAP dataset. The average classification rates are shown in Table I. For arousal classification, batch normalization plays an essential role since it increases the classification rates by about 10% in the 2-D and 3-D architectures with a dense prediction layer. Meanwhile, applying dense prediction also shows a significant increase of about 11% in both 2-D and 3-D architectures with batch normalization. Furthermore, among all the configurations, the 2-D architecture without batch normalization and dense prediction performs the worst with classification rate of 55.2%. Critically, the 3-D CNN architecture with batch normalization and dense prediction achieves the highest classification rate of 73.1%. For valence classification, the same importance of

[1]https://github.com/WangYiOtago/EmotioNet

TABLE I
THE AVERAGE CLASSIFICATION RATES (%) OF ARCHITECTURES WITH DIFFERENT CONFIGURATIONS.

	2-D CNN				3-D CNN			
	without DP		with DP		without DP		with DP	
	without BN	with BN	without BN	with BN	without BN	with BN	without BN	with BN
Arousal	55.2	58.1	60.5	69.5	59.4	62.5	61.7	**73.1**
Valence	57.0	57.5	55.2	66.8	56.1	57.2	56.6	**72.1**

Abbreviations: BN = Batch Normalization, DP = Dense Prediction.

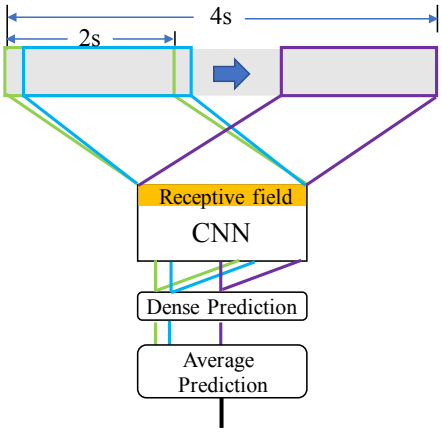

Fig. 8. The dense prediction layer.

TABLE II
A BRIEF COMPARISON BETWEEN OUR EXPERIMENTAL RESULTS AND PREVIOUS STUDIES

Papers	Accuracy(%)	
	Valence	Arousal
[22]	63.0	66.0
[23]	64.3	66.2
[19]	66.6	66.4
[20]	76.0	68.9
[21]	73.1	73.1
Average	68.6	68.1
2-D CNN	66.8	69.6
EmotioNet	72.1	73.1

Note: The results are ranked by the accuracy of arousal.

batch normalization and dense prediction is observed. Consequently, the best performance, i.e. 72.1%, is also obtained by the proposed 3-D CNN architecture with batch normalization and dense prediction.

According to our experimental results, we propose EmotioNet, a 3-D covariance shift adaptation-based CNN architecture with a dense prediction layer, which is shown in Fig. 7.

A brief comparison between our experimental results and the previous studies is shown in Table II. According to the table, comparing with the average result of the previous studies, the 2-D CNN, which is the classic EEG feature extraction scheme, achieved a relative lower accuracy in valence classification and a higher accuracy in arousal classification. As for the proposed EmotioNet, the valence and arousal classification rates increase 5.3% and 3.5% relative to the 2-CNN model. Comparing with previous studies, our arousal classification rate equals the best and our valence classification rate is higher than the average.

VI. DISCUSSION

According to these results, 3-D spatio-temporal architecture, batch normalization, and dense prediction are all essential techniques for our emotional EEG-specific CNN architecture. Specifically, the 3-D architecture improves the performance of the model by using a simultaneous temporal-spatial feature detection. As for batch normalization, although only a slight improvement was reported in an analysis of EEGNet [7], in our hands it significantly increases classification performance in the emotion recognition scenario. Dense prediction is also a crucial technique for emotion recognition since it enables the network to train on a long EEG window with fewer parameters, and so gather more general information in one training process and alleviate overfitting.

Unlike EEGNet and Braindecode, which use a feature extraction scheme that alternately extracts spatial and temporal EEG features, EmotioNet uses three 3-D convolution layers to extract the spatio-temporal emotional features and achieves higher classification rates.

Although EmotioNet achieves similar accuracy to previous methods, it has a clear advantage in not requiring handcrafted features. Its capability of automatic emotional feature detection provides a new method for EEG-based emotion recognition.

VII. CONCLUSION

In this paper, we propose EmotioNet, an emotional EEG-specific 3-D CNN architecture. Experimental results show that our model is capable of accurately recognizing emotions. We also show that 3-D architecture, batch normalization, and

dense prediction are critical techniques for an emotion recognition network. Furthermore, the idea of automatic emotional feature detection also provides a feasible way for emotion recognition in clinical and practical applications.

REFERENCES

[1] W. Heller and J. B. Nitscke, "Regional brain activity in emotion: A framework for understanding cognition in depresion," *Cognition & Emotion*, vol. 11, no. 5-6, pp. 637–661, 1997.

[2] K. Nashiro, M. Sakaki, and M. Mather, "Age differences in brain activity during emotion processing: Reflections of age-related decline or increased emotion regulation," *Gerontology*, vol. 58, no. 2, pp. 156–163, 2012.

[3] S. Koelstra, C. Muhl, M. Soleymani, J. S. Lee, A. Yazdani, T. Ebrahimi, T. Pun, A. Nijholt, and I. Patras, "DEAP: A database for emotion analysis using physiological signals," *IEEE Transactions on Affective Computing*, vol. 3, no. 1, pp. 18–31, 2012.

[4] Y. P. Lin, C. H. Wang, T. P. Jung, T. L. Wu, S. K. Jeng, J. R. Duann, and J. H. Chen, "EEG-based emotion recognition in music listening," *IEEE Transactions on Biomedical Engineering*, vol. 57, no. 7, pp. 1798–1806, 2010.

[5] P. C. Petrantonakis and L. J. Hadjileontiadis, "Emotion recognition from EEG using higher order crossings," *IEEE Transactions on Information Technology in Biomedicine*, vol. 14, no. 2, pp. 186–197, 2010.

[6] R. Jenke, A. Peer, and M. Buss, "Feature extraction and selection for emotion recognition from EEG," *IEEE Transactions on Affective Computing*, vol. 5, no. 3, pp. 327–339, 2014.

[7] V. J. Lawhern, A. J. Solon, N. R. Waytowich, S. M. Gordon, C. P. Hung, and B. J. Lance, "EEGNet: A Compact Convolutional Network for EEG-based Brain-Computer Interfaces," *arXiv preprint arXiv:1611.08024*, 2016.

[8] R. T. Schirrmeister, J. T. Springenberg, L. D. J. Fiederer, M. Glasstetter, K. Eggensperger, M. Tangermann, F. Hutter, W. Burgard, and T. Ball, "Deep learning with convolutional neural networks for EEG decoding and visualization," *Human brain mapping*, vol. 38, no. 11, pp. 5391–5420, 2017.

[9] G. Northoff, T. Witze, A. Richter, M. Gessner, F. Schlagenhauf, J. Fell, F. Baumgart, T. Kaulisch, C. Tempelmann, A. Heinzel *et al.*, "Gabaergic modulation of prefrontal spatio-temporal activation pattern during emotional processing: a combined fmri/meg study with placebo and lorazepam," *Journal of cognitive neuroscience*, vol. 14, no. 3, pp. 348–370, 2002.

[10] Y. Liu, O. Sourina, and M. K. Nguyen, "Real-time EEG-based human emotion recognition and visualization," in *2010 International Conference on Cyberworlds (CW)*. IEEE, 2010, pp. 262–269.

[11] S. Jirayucharoensak, S. Pan-Ngum, and P. Israsena, "EEG-based emotion recognition using deep learning network with principal component based covariate shift adaptation," *The Scientific World Journal*, vol. 2014, 2014.

[12] S. Blanco, H. Garcia, R. Q. Quiroga, L. Romanelli, and O. Rosso, "Stationarity of the EEG series," *IEEE Engineering in medicine and biology Magazine*, vol. 14, no. 4, pp. 395–399, 1995.

[13] A. I. Klistorner and S. L. Graham, "Electroencephalogram-based scaling of multifocal visual evoked potentials: effect on intersubject amplitude variability," *Investigative ophthalmology & visual science*, vol. 42, no. 9, pp. 2145–2152, 2001.

[14] S. Ioffe and C. Szegedy, "Batch normalization: Accelerating deep network training by reducing internal covariate shift," in *International conference on machine learning*, 2015, pp. 448–456.

[15] H. Gunes and B. Schuller, "Categorical and dimensional affect analysis in continuous input: Current trends and future directions," *Image and Vision Computing*, vol. 31, no. 2, pp. 120–136, 2013.

[16] 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.

[17] M. M. Bradley and P. J. Lang, "Measuring emotion: the self-assessment manikin and the semantic differential," *Journal of behavior therapy and experimental psychiatry*, vol. 25, no. 1, pp. 49–59, 1994.

[18] J. A. Russell, "A circumplex model of affect." *Journal of personality and social psychology*, vol. 39, no. 6, p. 1161, 1980.

[19] S. Y. Chung and H. J. Yoon, "Affective classification using bayesian classifier and supervised learning," in *2012 12th International Conference on Control, Automation and Systems (ICCAS)*. IEEE, 2012, pp. 1768–1771.

[20] V. Rozgić, S. N. Vitaladevuni, and R. Prasad, "Robust EEG emotion classification using segment level decision fusion," in *2013 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE, 2013, pp. 1286–1290.

[21] J. Atkinson and D. Campos, "Improving BCI-based emotion recognition by combining EEG feature selection and kernel classifiers," *Expert Systems with Applications*, vol. 47, pp. 35–41, 2016.

[22] R. Gupta and T. H. Falk, "Affective state characterization based on electroencephalography graph-theoretic features," in *2015 7th International IEEE/EMBS Conference on Neural Engineering (NER)*. IEEE, 2015, pp. 577–580.

[23] D. S. Naser and G. Saha, "Recognition of emotions induced by music videos using DT-CWPT," in *2013 Indian Conference on Medical Informatics and Telemedicine (ICMIT)*. IEEE, 2013, pp. 53–57.

[24] H. Candra, M. Yuwono, R. Chai, A. Handojoseno, I. Elamvazuthi, H. T. Nguyen, and S. Su, "Investigation of window size in classification of EEG-emotion signal with wavelet entropy and support vector machine," in *2015 37th Annual International Conference of the IEEE on Engineering in Medicine and Biology Society (EMBC)*. IEEE, 2015, pp. 7250–7253.