CONVOLUTIONAL NEURAL NETWORKS AND MULTITASK STRATEGIES FOR SEMANTIC MAPPING OF NATURAL LANGUAGE INPUT TO A STRUCTURED DATABASE

Mandy Korpusik, James Glass

MIT Computer Science and Artificial Intelligence Laboratory, Cambridge, MA 02139, USA
{korpusik, glass}@mit.edu

ABSTRACT

In this work, we investigate mapping both natural language food *and* quantity descriptions to matching USDA database entries. We demonstrate that a convolutional neural network (CNN) model with a softmax layer on top to directly predict the most likely database matches outperforms our previous state-of-the-art approach of learning binary classification and subsequently ranking database entries using similarity scores with the learned embeddings. The softmax model achieves 97.3% top-5 USDA quantity and 91.1% food recall over the full database, compared to only 70.0% quantity and 46.4% food recall with a sigmoid model, where top-5 recall indicates the percentage of test cases in which the correct quantity or food is in the top-5 hits. Evaluated on 9,600 spoken meals over all foods, the softmax model achieves 91.6% top-5 quantity and 80.1% food recall. We also explore jointly learning both mappings with a single CNN to boost quantity mapping, and improve food mapping by reranking the food database entries using the predicted quantity matches.

***Index Terms*—** Convolutional Neural Networks, Multitask Learning, Crowdsourcing, Semantic Embeddings, Reranking

1. INTRODUCTION

Today many Americans are tracking their diet, often to lose weight or to monitor specific nutrients, such as glucose levels for diabetics or sodium intake for those with high blood pressure. However, existing diet tracking applications can be too time-consuming for many users, requiring manually entering each eaten food one at a time and scrolling through a long list of potential database matches. Our proposed solution is a diet tracking application that uses speech and language understanding technology to enable quick, intuitive diet tracking; that is, a user simply speaks or types a natural language description of their meal, and our technology automatically determines the most likely food database matches [1, 2, 3].

In our prior work, we investigated the problem of mapping natural language meal descriptions to their corresponding food database entries. But this was limited to food matching, whereas we also need to address the remaining challenge of mapping user-described quantities to matching database quantity entries. This is a difficult problem because user descriptions are often very different from database entries. For example, a user might say "a bowl" or "a handful," but these do not easily map to database quantities, such as cups or grams. In a scenario where the user says, "**a spoonful** of *peanut butter*," the system should determine that the database food match is *Peanut butter, smooth style, with salt* with the corresponding quantity **1 tbsp**.

This research was sponsored by a grant from Quanta Computing, Inc., and by the Department of Defense (DoD) through the National Defense Science Engineering Graduate Fellowship (NDSEG) Program.

Meal	# Quantities	# Foods	# Diaries
Breakfast	616	1,477	33,317
Dinner	613	2,556	23,094
Salad	173	232	2,446
Sandwich	234	372	4,474
Smoothies	214	382	5,789
Pasta/Rice	366	1,262	12,715
Snacks	725	1,334	12,041
Fast Food	271	661	5,474
All Foods	1,562	5,156	99,350

Table 1. AMT data statistics, organized by meal.

In this paper, we tackle the quantity mapping problem by developing a new convolutional neural network (CNN) architecture that is trained with a softmax layer on top to directly predict the most likely database quantities, whereas our prior food mapping work used a binary classification network to learn embeddings for each database food entry, which were then ranked via cosine similarities at test time. In addition, we explore multitask learning to jointly predict both the matching food and quantity database entries given a single input meal description. We show that by leveraging the close relationship between quantities and foods, we can use predicted quantity matches to improve food ranking performance.

The remainder of the paper is organized as follows. First we describe the data collection process for obtaining natural language quantity descriptions and matching database entries. We then discuss the CNN architectures we explored and the multitask learning paradigm, followed by experimental results and discussion. Finally, we review related work and conclude with directions for future work.

2. DATA COLLECTION

Previously [4], we collected 31,712 meal descriptions and associated USDA food database matches via crowd-sourcing with Amazon Mechanical Turk (AMT). In order to generate intuitive meal description tasks, we partitioned the 5,156 database foods into eight meal categories, such as breakfast and dinner (see Table 1), and collected over 99k food and quantity descriptions in total. To collect quantity descriptions for our new work, we revised the AMT task such that workers were told to select one quantity option from among all the database quantity units available for a given food item. Then they were instructed to describe this quantity naturally (e.g., *two cups of*), and in a separate textbox, to describe the food item (e.g., *chopped kale*). To reduce biasing the language used by workers, we included images of the food items along with the less natural USDA titles.

For our evaluation on speech data, we collected 9,600 spoken

meal descriptions on AMT (1,200 for each of the eight meal categories), using the Google Chrome speech recognizer. The data was collected the same way as the text data, but with speech instead of text, and as a single description for each combined food and quantity.

3. METHOD

We implemented two variants of the CNN architecture for mapping natural language quantity descriptions to the USDA database: the first is reminiscent of our prior work on food mapping [5], learning USDA quantity embeddings via binary classification with a sigmoid output, and the other is a new approach that directly predicts the most likely database matches via a softmax layer on top. In this section, we first describe two baseline methods for ranking database quantities (using the longest common substring, and number of exact token matches). We then detail our two sigmoid and softmax CNN models. Finally, we explain the new multitask training mechanism.

3.1. Baselines

A simple lexical approach for ranking the most likely database quantity entries, given a user's meal description, is to use the number of tokens[1] that are an exact match between the two. Those database quantities with the maximum number of tokens in common would be ranked most highly. Our second baseline uses the length of the longest common substring (LCS) between the user's meal description and each database quantity, where we implement a string matching algorithm that stores the number of matching characters seen so far in a dynamic programming table. For the food mapping task, we compare against our prior state-of-the-art CNN with reranking [5].

3.2. Sigmoid CNN

The sigmoid model (see Fig. 1) is the same as that used in our previous work [4] for mapping natural language meal descriptions to their associated food database matches, except we pad the input quantity descriptions to 20 tokens instead of 100.[2] The input 50-dimension embedding layer is followed by a 1D convolution with 64 filters spanning windows of three tokens, with a rectified linear unit (ReLU) activation and dropout of probability 0.2. This network is trained for a binary verification task, where each input pair consists of a user-described quantity and a USDA quantity that either matches the user's description or not.[3] Through learning to complete this binary verification task, the network learns semantic embedding representations of each USDA database quantity, which are then used at test time to rank all the possible database quantity options based on the cosine similarity score with the user-described quantity embedding (which is generated by feeding the input meal description through the meal portion of the network, consisting of an embedding layer followed by a convolution and max-pooling). The model is trained with the Adam optimizer [6] on binary cross-entropy loss.

3.3. Softmax CNN

The softmax CNN (see Fig. 2) is a new architecture that we implemented to directly rank all the USDA database quantities within the network itself, rather than requiring the multi-step process of generating embeddings with the network and subsequently ranking all the

[1]We used the Spacy toolkit (https://spacy.io/) for tokenization.

[2]The padding results in dot products with each of the 20 input tokens.

[3]For each positive match we collected, we sampled a random negative quantity from among those quantities not described by the user.

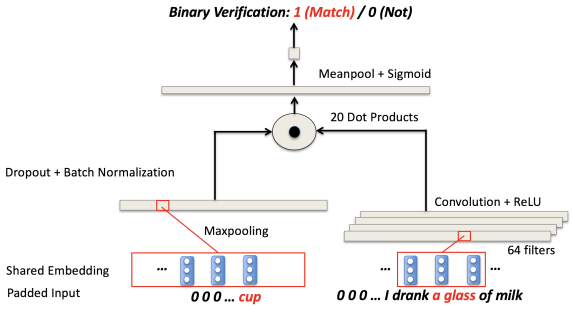

Fig. 1. The sigmoid CNN for predicting whether or not an input USDA quantity name and user-described quantity match or not.

USDA quantities with cosine similarity scores. Rather than feeding the network only a single USDA quantity option, we input all possible USDA quantities along with the user's meal description. The USDA quantities are embedded and used in dot product computation with the convolved meal description the same way as in the sigmoid network; however, this model performs the computation for *every* USDA quantity, with a final feed-forward layer on top that outputs a probability distribution over all the quantities via a softmax.

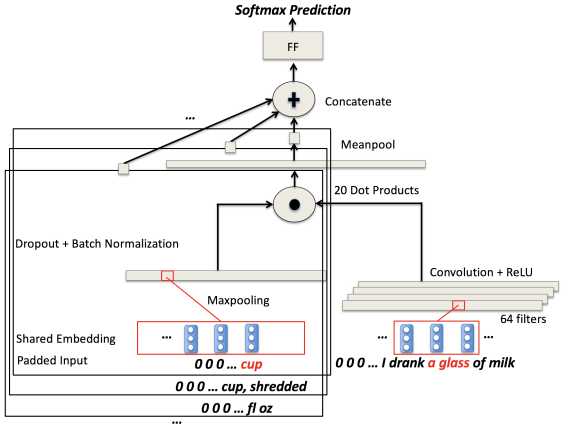

Fig. 2. The softmax CNN for directly ranking all the USDA quantity options for a given user's input meal description.

3.4. Simple Softmax CNN

The simple softmax CNN (see Fig. 3) is another new neural architecture that feeds only the input meal description into the embedding and convolution layers before the final feed-forward layer with a softmax output over all possible food or quantity options.

3.5. Multitask CNN

The new multitask model is structured the same way as the sigmoid and softmax CNNs for quantity mapping, but has an additional output layer for predicting the USDA food match. Thus, the majority of the network is shared between the two tasks, and the loss is the combination of the quantity prediction and food prediction losses.

4. EXPERIMENTS

Here, we demonstrate that the new softmax model outperforms the state-of-the-art CNN [5] that we used previously for food mapping.

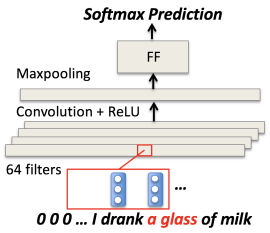

Fig. 3. The simple softmax CNN for directly ranking all the USDA quantity options for a given user's input meal description.

We also show that jointly training the CNN to predict both USDA food and quantity matches yields higher quantity recall for most meal categories. We note that since the quantity predictor has high performance, we can leverage the predicted quantities to rerank the USDA food options to favor those that have the highest ranked quantities as available options in the database, which consistently boosts performance. We evaluate on both written and spoken held-out test sets[4] using top-5 recall, which indicates the percentage of test cases in which the correct food or quantity option appears in the top-5 hits.

4.1. Sigmoid vs. Softmax

We can see in Table 2 that all the CNN models significantly outperform the longest common substring (LCS) and number of word matches (WM) baselines on the quantity mapping task; thus, simply counting word or character matches is not sufficient, and a more sophisticated model is required. The softmax is superior to the sigmoid, for both food and quantity matching (trained separately). For all subsequent experiments, we use the simple softmax as the default softmax since it performs best; we conjecture this is due to the relatively simple task, for which the complex softmax is too powerful and overfits the training data. In Table 3, we evaluate the new CNN models on food mapping and discover that the simple softmax model improves our prior state-of-the-art CNN (using a word-by-word similarity reranking algorithm [5]) with an 83.3% gain on all foods.

Meal	LCS	WM	Sigm.	Soft.	Simple Soft.
Breakfast	13.5	7.87	71.1	94.8	**96.9**
Dinner	10.9	10.4	82.1	94.8	**98.2**
Salad	25.1	36.9	75.5	82.7	**97.4**
Sandwich	19.2	30.1	77.7	92.0	**97.2**
Smoothies	18.5	37.1	75.3	92.6	**98.7**
Pasta/Rice	11.7	12.6	84.0	95.5	**98.1**
Snacks	15.8	12.3	63.7	93.4	**96.9**
Fast Food	16.5	13.7	72.2	93.8	**98.7**
All Foods	13.9	13.3	70.0	96.5	**97.3**

Table 2. Top-5 quantity recall per meal category for LCS and word match (WM) baselines, the sigmoid, softmax, and simple softmax.

4.2. Spoken Data

Because our production system will enable both text and speech input, here we investigate whether the models trained on text data still perform well on speech data. As shown in Table 4, when evaluated on 9,600 spoken meal descriptions (1,200 per meal category), the softmax quantity and food mapping models still perform quite well.

[4]We divide the 99,350 text samples into 80% train/10% dev/10% test.

Meal	Baseline CNN	Sigmoid	Simple Soft.
Breakfast	47.3	34.6	**95.8**
Dinner	38.5	25.4	**91.6**
Salad	75.9	40.4	**98.4**
Sandwich	70.8	46.4	**97.9**
Smoothies	69.5	53.7	**97.2**
Pasta/Rice	39.2	29.9	**89.5**
Snacks	60.2	41.4	**96.9**
Fast Food	53.6	38.7	**98.2**
All Foods	49.7	46.4	**91.1**

Table 3. Top-5 food recall per meal category for the new sigmoid and softmax CNNs, compared to the baseline CNN reranker [5].

Meal	Q LCS	Q WM	Q Best	F CNN	F Best
Breakfast	15.2	8.36	92.0	54.5	80.9
Dinner	14.5	14.6	92.8	45.7	67.3
Salad	27.2	40.9	89.5	82.6	90.9
Sandwich	24.8	29.3	89.2	82.4	88.8
Smoothies	26.0	37.8	90.8	73.6	88.6
Pasta/Rice	14.5	14.0	89.4	43.7	60.2
Snacks	19.5	16.8	90.3	64.4	84.1
Fast Food	19.0	13.3	84.3	58.2	84.0
All Foods	20.1	21.9	91.6	62.6	80.1

Table 4. Top-5 quantity (Q) and food (F) recall per meal category for the best simple softmax and baseline models on *spoken* data.

4.3. Quantity Input Only

Since it would seem that the interaction between foods and quantities helps the models learn to predict relevant foods and quantities, we ran an experiment to see whether the quantity mapping performance would suffer if the input to the network was only the quantity description, without the associated food's description. For example, with the user-described input meal diary "I had a cup of cheese," the model might tend to prefer database units that relate to cheese, such as "cup, shredded" rather than the generic "cup" or "cup, diced." However, in this experiment, the input would simply be "cup." To convert the full meal to a quantity segment, we ran our CNN tagger from prior work [3] that labels food and quantity tokens, and extracted only the tagged quantity tokens. As expected, quantity mapping performance is much worse without the full meal input. On Breakfast, the top-5 quantity recall is only 55.8 for sigmoid and 70.0 for softmax (leading to 21.5% and 26.2% drop in performance for the sigmoid and softmax models, respectively, with only quantities as input); the top-5 quantity recall scores for the other meals are similarly all below 70 for sigmoid models and below 80 for softmax.

4.4. Multitask Learning

Finally, we investigated multitask learning (MTL) to determine whether a single model that jointly predicts food and quantity labels would perform better than either model individually. MTL with the simple softmax model improves quantity mapping for most meal categories (see Table 5); however, the food mapping task is more challenging, as there are far more food options than quantities, so MTL does not benefit this task. This indicates that MTL can improve the task with fewer labels, but not the more challenging task [7].

Since we also want to boost food mapping by leveraging the

quantity mapping task, as an alternative approach to training a joint multitask model, we used the best quantity softmax trained on all data (since if we only used training data, then it could not accurately predict quantities seen only in the test data) to rerank the predicted foods. This boosts the top-1 food mapping performance on test data for all meals except Fast Food (see Table 5). First, we predict the top-5 USDA quantities. Then, we rerank the predicted USDA foods that have at least one of the top-5 predicted quantities as a unit option above those that do not. This gain indicates that we can leverage a higher-performing task to improve a weaker, closely related task.

Meal	Q Soft.	MTL Q	F Soft.	Reranked F
Breakfast	**89.6**	88.7	80.4	**81.5**
Dinner	89.6	**89.7**	71.8	**72.9**
Salad	**89.0**	88.2	83.7	**84.1**
Sandwich	**89.6**	88.4	82.7	**83.6**
Smoothies	89.3	**89.9**	82.0	**82.0**
Pasta/Rice	90.0	**90.1**	66.5	**67.4**
Snacks	85.9	**86.5**	82.4	**83.0**
Fast Food	92.5	**93.4**	**88.8**	88.3
All Foods	**86.6**	86.3	70.3	**71.1**

Table 5. Top-1 recall for MTL Quantity and reranked Foods.

5. DISCUSSION

When users interact with our live nutrition system, we must ensure the rankings generated by our food and quantity mappers at test time are reasonable. To qualitatively evaluate the performance of our CNN model, we observe that its predictions make sense intuitively. For example, in the test meal description "*I had a **cup** of milk and a **tablespoon** of honey*," with the softmax model trained on Breakfast data, the quantity ranking for milk is {cup, fl oz, quart} and {tbsp, cup, packet (0.5 oz)} for honey, which matches commonsense.[5]

By inspecting the nearest neighbors of the learned USDA quantity embeddings (see Table 6), we see that the Pasta Softmax model (i.e., the complex softmax CNN trained on the Pasta meal category[6]) is learning meaningful semantic representations of quantities, where those of a similar unit are close to each other in vector space. We can also determine what the 64 CNN filters over the embedded quantities learned by inspecting which tokens cause the filters to fire with the highest activations. This analysis shows that filter 46 tends to identify meat-related tokens (i.e., tenderloin, beef, loin, strip, steak, pork, wagyu, roast, dried, and strips are the top-10 tokens in order of descending filter response), while filter 53 picks out numbers (i.e., three, one, a, two, eight, five, four, six, twelve, and seven).

Quantity	Neighbor 1	Neighbor 2	Neighbor 3
cup	cup whole	cup slices	cup shredded
oz	oz whole	oz boneless	oz serving 2.7 oz
serving 1/2 cup	serving 1 cup	cup slices	cup whole

Table 6. Top-3 neighbors to three USDA quantities, based on Euclidean distance of learned embeddings from a Pasta softmax model.

6. RELATED WORK

Multitask learning (MTL) has been applied successfully to many natural language processing (NLP) tasks. Collobert *et al.*'s early exploration of multitask learning involved jointly training a single CNN like ours on several tasks: part-of-speech tagging, chunking, named entity recognition, semantic role labeling (SRL), semantic relation prediction, and language modeling (LM) [8]. They focus specifically on SRL, while we care about both our tasks equally. Liu *et al.* built a multitask deep neural network (DNN) that combined two different tasks of multiple-domain query classification and information retrieval for web search ranking [9]. Similar to our work, they embedded an input query into a lower-level shared semantic representation used for the two different tasks at the top layer; however, they use a DNN while we employ a CNN.

Other work in MTL for NLP demonstrated an improvement in sentence compression by incorporating two auxiliary tasks, combinatory categorical grammar (CCG) tagging and gaze prediction, based on the intuition that longer reading time correlates with text difficulty [10]; they showed that the cascaded architecture, where auxiliary tasks are predicted at an inner layer, outperforms the model where auxiliary tasks are predicted at the top layer. Luong *et al.* investigated MTL for neural machine translation with the sequence-to-sequence model, with the surprising result that parsing (i.e., sharing the encoder) and image caption generation (i.e., sharing the decoder) both improve translation, despite the much smaller datasets [11].

Multi-task learning has also been applied to other fields, including speech recognition and computer vision. Toshniwal *et al.* explored end-to-end speech recognition on the conversational Switchboard corpus, demonstrating gains in character-based automatic speech recognition (ASR) by adding supervision at lower layers in a deep long short-term memory (LSTM) network with two lower-level tasks [12]. In computer vision, Misra *et al.* proposed a novel cross-stitch unit that combines CNNs for two tasks by automatically learning an optimal combination of shared and task-specific representations [13]. In addition, Wang *et al.* constructed a shared sub-network with higher-level sub-networks for two image representations, in order to achieve high accuracy from cross-image representations while maintaining the efficiency of single-image representations [14].

Another area of work related to ours is that of learning joint embeddings. Prior work used a margin-based hinge loss to rank annotations given an image [15], learned a joint multimodal space between images and captions for caption generation [16, 17], and learned sentence or document embeddings [18]. Recently, CNNs have also gained popularity among the NLP community, achieving state-of-the art performance on text classification [19, 20, 21]. Finally, parallel CNNs [22, 23, 24], attention-based CNN (ABCNN) models [25], and hierarchical ABCNNs [26] have been proposed for sentence matching and machine comprehension.

7. CONCLUSION AND FUTURE WORK

In this paper, we expanded our prior work mapping natural language meal descriptions to their corresponding USDA *food* database entries to address the remaining challenge of mapping meal descriptions to their associated *quantity* database hits. We have shown that a new softmax CNN model outperforms our previous best sigmoid CNN trained on a binary verification task, and achieves 91.6% top-5 quantity recall on a spoken test set of 9,600 meal descriptions over the full USDA database. We investigated multitask learning to improve quantity mapping, and demonstrated that we can leverage the high recall of the quantity predictor to improve food ranking. In future work, we will investigate contextual understanding to determine whether the user has refined their meal description, and run a pilot study with nutritionists' patients. We may explore speech-to-speech networks and input lattices to account for speech recognition errors.

[5] A pre-trained semantic tagger [3] identifies each food/quantity segment.

[6] In the deployed system, we would use the full Allfood Softmax model.

8. REFERENCES

[1] M. Korpusik, N. Schmidt, J. Drexler, S. Cyphers, and J. Glass, "Data collection and language understanding of food descriptions," *Proceedings of 2014 IEEE Spoken Language Technology Workshop (SLT)*, 2014.

[2] M. Korpusik, C. Huang, M. Price, and J. Glass, "Distributional semantics for understanding spoken meal descriptions," *Proceedings of 2016 IEEE Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2016.

[3] M. Korpusik and J. Glass, "Spoken language understanding for a nutrition dialogue system," *IEEE Transactions on Audio, Speech, and Language Processing*, 2017.

[4] M. Korpusik, Z. Collins, and J. Glass, "Semantic mapping of natural language input to database entries via convolutional neural networks," *Proceedings of IEEE Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2017.

[5] M. Korpusik, Z. Collins, and J. Glass, "Character-based embedding models and reranking strategies for understanding natural language meal descriptions," *Proceedings of Interspeech*, 2017.

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

[7] J. Bingel and A. Søgaard, "Identifying beneficial task relations for multi-task learning in deep neural networks," in *Proceedings of the 15th Conference of the European Chapter of the Association for Computational Linguistics: Volume 2, Short Papers*, 2017, pp. 164–169.

[8] R. Collobert and J. Weston, "A unified architecture for natural language processing: Deep neural networks with multitask learning," in *Proceedings of the 25th International Conference on Machine Learning (ICML)*. ACM, 2008, pp. 160–167.

[9] X. Liu, J. Gao, X. He, L. Deng, K. Duh, and Y. Wang, "Representation learning using multi-task deep neural networks for semantic classification and information retrieval," in *Proceedings of the Human Language Technologies: The 2015 Annual Conference of the North American Chapter of the ACL (HLT-NAACL)*, 2015, pp. 912–921.

[10] S. Klerke, Y. Goldberg, and A. Søgaard, "Improving sentence compression by learning to predict gaze," in *Proceedings of the Human Language Technologies: The 2016 Annual Conference of the North American Chapter of the ACL (HLT-NAACL)*, 2016, pp. 1528–1533.

[11] M. Luong, Q. Le, I. Sutskever, O. Vinyals, and L. Kaiser, "Multi-task sequence to sequence learning," in *Proceedings of the International Conference on Learning Representations (ICLR)*, 2016.

[12] S. Toshniwal, H. Tang, L. Lu, and K. Livescu, "Multitask learning with low-level auxiliary tasks for encoder-decoder based speech recognition," *arXiv preprint arXiv:1704.01631*, 2017.

[13] I. Misra, A. Shrivastava, A. Gupta, and M. Hebert, "Cross-stitch networks for multi-task learning," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2016, pp. 3994–4003.

[14] F. Wang, W. Zuo, L. Lin, D. Zhang, and L. Zhang, "Joint learning of single-image and cross-image representations for person re-identification," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2016, pp. 1288–1296.

[15] J. Weston, S. Bengio, and N. Usunier, "Wsabie: Scaling up to large vocabulary image annotation," in *Proceedings of the Twenty-Second international joint conference on Artificial Intelligence (IJCAI)*, 2011, vol. 11, pp. 2764–2770.

[16] A. Karpathy and L. Fei-Fei, "Deep visual-semantic alignments for generating image descriptions," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2015, pp. 3128–3137.

[17] D. Harwath, A. Torralba, and J. Glass, "Unsupervised learning of spoken language with visual context," in *Proceedings of the 30th Conference on Neural Information Processing Systems (NIPS)*, 2016, pp. 1858–1866.

[18] K. Hermann and P. Blunsom, "Multilingual models for compositional distributed semantics," in *Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics (ACL)*, 2014.

[19] A. Conneau, H. Schwenk, Y. Lecun, and L. Barrault, "Very deep convolutional networks for text classification," in *Proceedings of the 15th Conference of the European Chapter of the Association for Computational Linguistics: Volume 1, Long Papers*, 2017, p. 11071116.

[20] X. Zhang, J. Zhao, and Y. LeCun, "Character-level convolutional networks for text classification," in *Proceedings of Advances in neural information processing systems (NIPS)*, 2015, pp. 649–657.

[21] Y. Xiao and K. Cho, "Efficient character-level document classification by combining convolution and recurrent layers," in *Proceedings of the Thirtieth International Florida Artificial Intelligence Research Society Conference*, 2017, pp. 353–358.

[22] L. Pang, Y. Lan, J. Guo, J. Xu, S. Wan, and X. Cheng, "Text matching as image recognition," in *Proceedings of the Thirtieth AAAI Conference on Artificial Intelligence*, 2016, pp. 2793–2799.

[23] Z. Wang, H. Mi, and A. Ittycheriah, "Sentence similarity learning by lexical decomposition and composition," in *Proceedings of COLING 2016, the 26th International Conference on Computational Linguistics: Technical Papers*, 2016, pp. 1340–1349.

[24] B. Hu, Z. Lu, H. Li, and Q. Chen, "Convolutional neural network architectures for matching natural language sentences," in *Proceedings of Advances in neural information processing systems (NIPS)*, 2014, pp. 2042–2050.

[25] W. Yin, H. Schütze, B. Xiang, and B. Zhou, "ABCNN: Attention-based convolutional neural network for modeling sentence pairs," in *Transactions of the Association for Computational Linguistics*, 2016, vol. 4, pp. 259–272.

[26] W. Yin, S. Ebert, and H. Schütze, "Attention-based convolutional neural network for machine comprehension," in *Proceedings of 2016 NAACL Human-Computer Question Answering Workshop*, 2016, pp. 15–21.