AFRL-AFOSR-JP-TR-2017-0025

Fusion and Sense Making of Heterogeneous Sensor Network and Other Sources

Zhongxiang Shen
NANYANG TECHNOLOGICAL UNIVERSITY

03/16/2017
Final Report

DISTRIBUTION A: Distribution approved for public release.

Air Force Research Laboratory
AF Office Of Scientific Research (AFOSR)/ IOA
Arlington, Virginia 22203
Air Force Materiel Command

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

1. REPORT DATE *(DD-MM-YYYY)*
16-03-2017

2. REPORT TYPE
Final

3. DATES COVERED *(From - To)*
28 Feb 2013 to 27 Mar 2016

4. TITLE AND SUBTITLE
Fusion and Sense Making of Heterogeneous Sensor Network and Other Sources

5a. CONTRACT NUMBER

5b. GRANT NUMBER
FA2386-13-1-4083

5c. PROGRAM ELEMENT NUMBER
61102F

6. AUTHOR(S)
Zhongxiang Shen

5d. PROJECT NUMBER

5e. TASK NUMBER

5f. WORK UNIT NUMBER

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)
NANYANG TECHNOLOGICAL UNIVERSITY
50 NANYANG AVENUE
SINGAPORE, 639798 SG

8. PERFORMING ORGANIZATION REPORT NUMBER

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)
AOARD
UNIT 45002
APO AP 96338-5002

10. SPONSOR/MONITOR'S ACRONYM(S)
AFRL/AFOSR IOA

11. SPONSOR/MONITOR'S REPORT NUMBER(S)
AFRL-AFOSR-JP-TR-2017-0025

12. DISTRIBUTION/AVAILABILITY STATEMENT
A DISTRIBUTION UNLIMITED: PB Public Release

13. SUPPLEMENTARY NOTES

14. ABSTRACT
PI Team has proposed to utilize an adaptive multimodal fusion framework that uses both training data and web resources for scene classification, the experimental results on the benchmark datasets show that the proposed text-aided scene classification framework could significantly improve classification performance. Experimental results also show that the adaptive multimodal fusion mechanism can effectively individualize reliability-dependent weighting modulation for every new observation.

15. SUBJECT TERMS
Multimodality, Network Congestion;, Network Vulnerability, Sensemaking

16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON
a. REPORT	b. ABSTRACT	c. THIS PAGE			HONG, SENG
Unclassified	Unclassified	Unclassified	SAR	13	19b. TELEPHONE NUMBER *(Include area code)* 315-229-3519

Standard Form 298 (Rev. 8/98)
Prescribed by ANSI Std. Z39.18

Final Report for AOARD Grant FA2386-13-1-4083

"Fusion and Sense Making of Heterogeneous Sensor Networks and Other Sources"

Date January 19, 2017

Name of Principal Investigators (PI and Co-PIs): Shen Zhongxiang, Mao Kezhi
- e-mail address : ekzmao@ntu.edu.sg
- Institution: Nanyang Technological University
- Mailing address: Block S2.1, School of EEE, Nanyang Technological University, Singapore
- Phone: +65 6790 4284

Period of Performance: March/28/2013 – Sept/27/2016

Abstract:

To train a scene classifier with good generalization capability, a large number of human labeled training images are often needed. However, a large number of well-labeled training images may not always be available. To alleviate this problem, the web resources-aided scene classification framework was proposed. The present project is a new development based on our previously proposed framework, with the following improvements. First, a text-based filtering algorithm is developed to remove irrelevant web search returns since irrelevant web search returns provide irrelevant or even wrong information about the class of an image. Second, an adaptive fusion algorithm is developed for the integration of visual feature-based and web textual feature-based classification results. This adaptive fusion algorithm is inspired by the multisensory integration mechanism of human whose adaptability is achieved by reliability-dependent weighting of different sensory modalities. Experimental results show that the proposed web textual resources aided image classification framework can improve classification accuracy of some classes by 13% and 12% in the UIUC-Sports and LabelMe8 datasets, respectively.

Introduction:

As an important issue in visual recognition tasks, image classification has received considerable attentions. A supervised learning-based image classification system often demands a large number of labeled training images. However, a large number of training images are not always available, and even available, labeling of the training images is usually tedious and time-consuming. To relieve the shortage of labeled training data, seeking help from open resources on the World Wide Web has been proposed as a solution. In the literature, a few homogeneous web data (training data and web resources are in the same modality) aided image classification approaches have been proposed, including self-taught learning, domain adaptation, semi-supervised learning and etc. By taking the web images as normal training images, these approaches employ low-level image descriptors such as SIFT and GIST to extract visual features and then input the features to classifiers. The use of web data indeed relieves the shortage of training data, but these works only use the visual features of the web images, without using high-level semantic textual features carried by the tags or captions of the web images.

To address the above mentioned limitations, heterogeneous web data (such as text) aided framework has been explored. In this framework, task-dependent web databases containing images and text annotations downloaded from the web are constructed. By exploring the texts affiliated to web images, high-level semantic features reflecting image contents can be extracted. Usually, the visual features extracted from images and the textual features extracted from web text are used separately to train respective classifiers. The image and text modalities are fused on the decision level through linear combination of the classifiers, where the weights of the classifiers are trainable and are fixed once trained.

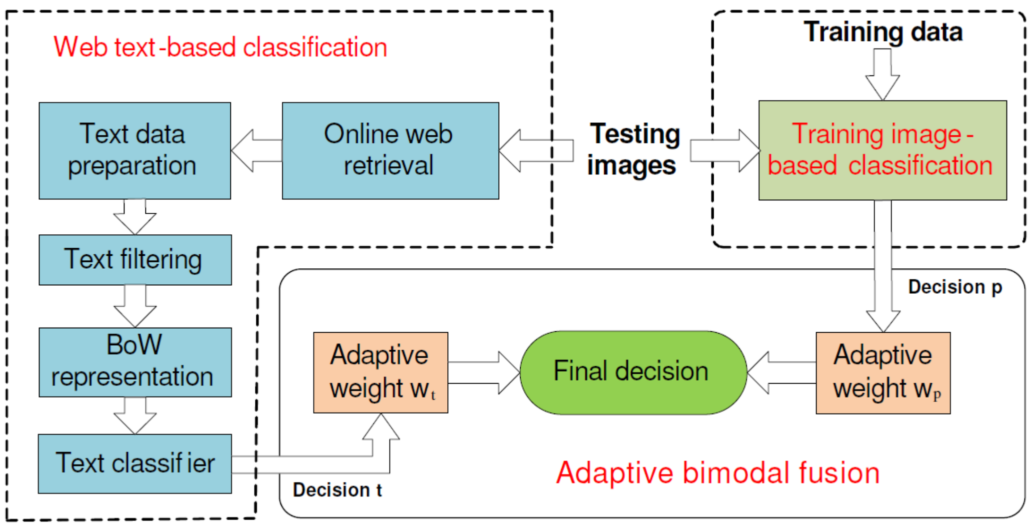

Fig. 1. An overview of the proposed framework with text filtering and adaptive multimodal fusion, which combines results of two classifiers using adaptive weighting.

In many existing works, the textual features of a testing image is extracted from the text affiliated to the visually similar images in the pre-constructed databases. To obtain good generalization capability of the textual feature extraction, the pre-constructed web databases need to be very large to have a good coverage of the testing data. To alleviate this requirement, we proposed an online search-based method for textual feature extraction and textual feature-based image classification. The online search-based method makes use of the powerful search engine of the Google reverse image search and regards all the resources on the World Wide Web as the task-independent databases. Thus, it is more likely to find visually similar images from the web resources. In addition, the online search-based method is more efficient for the extraction of textual data from web because it does not have to download web images. Since the search of visually similar web images for a testing image is not limited to the task-oriented pre-constructed databases, the online search-based method has the potential to provide semantic textual information to correctly classify images from unseen new classes. But one challenge encountered in the online search-based method for textual feature extraction is that the web resources are often noisy, and the returned texts of online web search contain inaccurate or even wrong information about the class of the query (testing) image. To deal with this problem, our previous work attempted to extract class label directly from the texts of the returned images. Simplicity is the advantage of this method, but it dismisses all other information except class labels. If the returned text does not contain the class label of the image, the entire text will be discarded, which is a kind of waste of information.

In this report, we improve upon our previous work. Besides image class labels, we attempt to extract other semantic information underlying the web text. To address this, web texts are represented and classified using the Bag-of-Words model in this project. To deal with the noise in the returned online search results, a filtering algorithm is developed to remove irrelevant web texts. By filtering, we can make full use the information within the web texts retained while stay away from the irrelevant information sources. Since web resources have great reliability diversity, it may not be an optimal practice to allocate fixed weights to the visual feature-based and textual feature-based classifiers. In this project, an adaptive fusion algorithm is developed for the integration of the visual feature-based and web textual feature-based classification results. This adaptive fusion algorithm is inspired by the multisensory integration mechanism of human whose adaptability is achieved by reliability-dependent weighting of different sensory modalities.

As shown in Fig. 1, our proposed framework consists of three components: training image-based classification, web text-based image classification, and an adaptive multimodal fusion. Two separate classifiers based on visual and textual features are first built, and the decision-level fusion is then performed by applying pairwise adaptive weight vectors w_t and w_p to image-based and text-based classification scores. The goal of handling different modality of data separately is to reduce the vulnerable interaction of heterogeneous data. In the following sections, details of the three components are presented respectively.

Proposed Methods:

Our previous work has demonstrated the effectiveness of using web resources to aid classification of images that are hard to classify based on a limited number of training data only. Simplicity is the advantage of this approach, however, it dismisses all other information underlying the text except class label. In other words, the limitation of the approach is that it discards the entire text if the class label information is not found from the text. In this report, we aims to overcome this problem by representing web text using Bag-of-Words (BoW) model and then inputting the vector representation of the text to a classifier to decide the class label of the image.

(1) Web Text-Based Image Classification:

Google reverse image search is employed in this study. In Google image search, an image is uploaded as a query to search visually similar images on the web. The Google image search returns a list of images sorted based on visual similarity. The web texts including image captions and descriptions of the first *n* returned images are then extracted. Since the similar images from web resources are already annotated, the web texts contain indicative semantic information about the class of the query image.

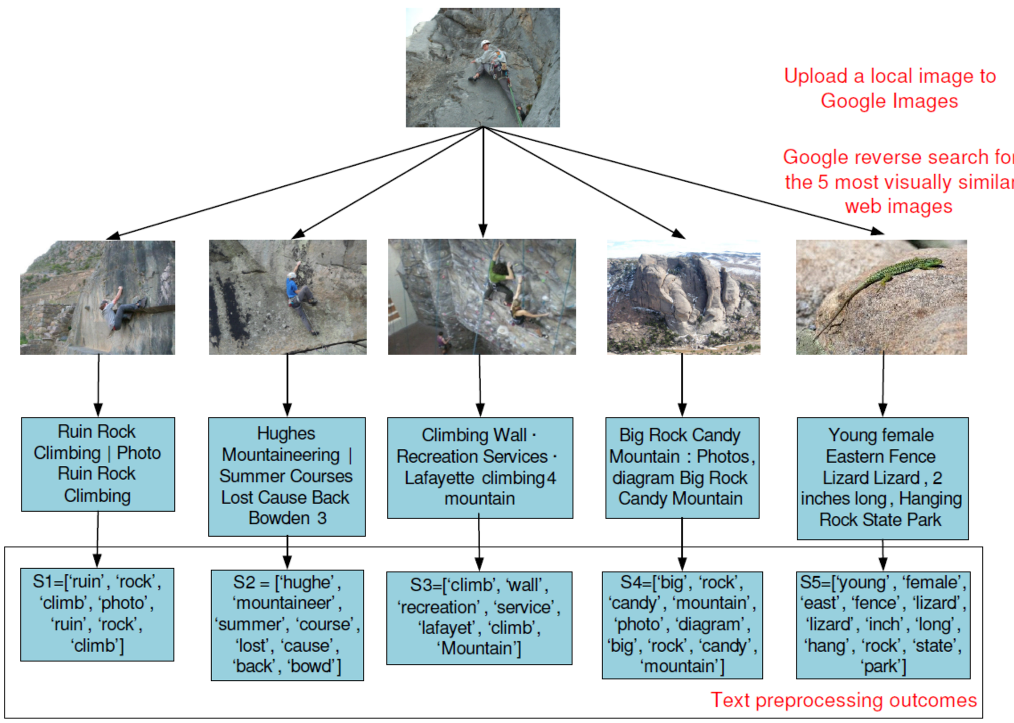

Fig. 2. An example of the procedure of on-line text retrieval results when the class label is "climbing" and $n = 5$. It is observed that the fifth image is a incorrect return "lizard" and its affiliated text information is certainly unrelated to the target image.

Fig. 2 illustrates the general steps of the text data preparation. Once the raw text is converted from a list of words to strings, lexicons are generated from the raw strings by lexical analysis, which is known as tokenization. Tokenization is followed by data cleaning such as predefined ``stopping words'' (pronouns, connectives, prepositions, etc.) removal and encoded data (ASCII, Lation-2, UTF-8, etc.) extraction. Next, morphological affixes are removed from words, which is called word stemming. For instance, given a raw text *Two dogs are chasing \u0061 boy on the road*, the outcome from text preprocessing becomes a string of ['two', 'dog', 'chase', 'boy', 'road'].

Exploring web resource automatically by machine itself is a very challenging task since web resources are often non-cooperative and noisy, for example, the returned images and texts might be irrelevant to the query image as shown in the fifth returned image and text in Fig. 2. On one hand, we hope to increase the volume of the web texts so that the retrieved web text resources are sufficiently abundant. But on the other hand, the irrelevant text should be discarded since irrelevant texts provide irrelevant or even wrong information about the query image.

Vector space model (VSM) is a popular method for text representation and text similarity evaluation. To remove the task-unrelated text, we measure the cosine similarity between the web returned text data and the relevant reference documents with respect to the class labels and related tags. Yet, one challenging issue here is that the returned web texts describing the web images are usually much shorter than the reference documents, and the short-to-long document similarity is hard to measure. To address this problem, we simultaneously vectorize the long and short documents and map the text vectors into a codebook, which is learned from a feature ranking algorithm upon the reference documents (i.e., after the documents are represented by the TF-IDF model, top 15% codes are selected by the chi-squared test). Given a test image, Google Images returns the textual modality of *n* visually similar web images as a document of *n* strings. Accordingly, the document is vectorized into a collection of *n* sub-vectors $s=[s_1, s_2, ..., s_n]$, and the web reference documents are vectorized into a set of *h* sub-vectors $d=[d_1, d_2, ..., d_h]$. Next, we evaluate the cosine similarity between each individual sub vectors s_i and d_j.

Using the text filtering algorithm above, we could effectively remove irrelevant noisy texts from the raw web text corpora. Next, a classifier is learned from the texts. In this report, BoW-based text representation is adopted, and the TF-IDF weighting feature vectors are computed for the web text collections of the query images. As the name implies, TF-IDF features are the combination of term frequency and inverse document frequency. Logarithmically scaled term frequency is commonly used. The TF-IDF features are the products of term frequency (tf) and inverse document frequency (idf), which reflect how important a word is to a document in a corpus.

The resulted TF-IDF feature vectors are then sent to a text classifier. Considering the multi-class issue of the training images, we implement one-vs-rest (one-vs-all) linear SVM. In this approach, *h* binary classifiers are employed, each of which separates class *j, j*=1, …, *h* from the rest *h-1* classes. Once the *h* (one-vs-rest) SVM classifiers have been trained, the TF-IDF feature vector of a testing image is supplied to the h classifiers. The resulting image-to-label vector of a query image is denoted by $t=[t_1, t_2, ..., t_h]^T$, where t_j denotes the decision score of the testing (query) image belonging to the j^{th} class.

(2) Training Image-based Classification:

A variety of visual feature extraction methods have been proposed in the literature. In this report, we investigate the performance of the BoW-based text classification combined with GIST, dense SIFT (PHOW), object bank and sparse coding Spatial Pyramid Matching (ScSPM) descriptors-based image classification respectively. GIST image descriptors are convolved with Gabor filters, where the returned feature maps are divided into grids to obtain the average sub-region feature values. PHOW features are extracted at multiple scales on multiple pyramid levels and quantized into the visual words using pyramid-based k-means clustering. Object bank employs many pre-defined object detectors such as latent Support Vector Machine (SVM) object detectors to obtain the responses by three dimensional spatial pyramid mapping. Non-negative ScSPM trains a codebook based on extracted SIFT features. The sparse coding features are then calculated by feature pooling using spatial pyramid.

After extracted from an image, the feature vector is input to a classifier. The outcome from the SVM classifier is a *h*-dimensional vector $p=[p_1, p_2, \ldots, p_h]^T$, where p_j denotes the decision score of this testing sample belonging to the j^{th} class. Here, the classification results refer to the decision scores. In general, the highest score $max_{1 \leq j \leq h}\{p_j\}$ corresponds to the correct class that the testing instance belongs to.

(3) Adaptive multimodal Fusion:

Human has superb adaptive capability. Research in neuron science, for example \cite{ohshiro2011normalization}, has found that the adaptive capability is owing to the multisensory integration mechanism through divisive normalization. Inspired by this finding in human multisensory integration, we propose an adaptive fusion algorithm at the decision level as shown in Fig. 1. In the adaptive fusion of the visual feature-based classifier and the textual feature-based classifier, the weights assigned to each class in each modality are adapted to each individual testing image based on the reliability of visual cues and textual cues of the testing image.

Multimodal fusion on the decision level often employs a weighted linear summation model. Different from the traditional multimodal fusion model that fixes the weights w_{pj} and w_{tj} for the visual feature-based and textual feature-based classifiers, we employ adaptive weights in this report. Weights w_{pj} and w_{tj} are now functions of coherence, normalized within the range 0 to 1.

Assuming there are *N* training data from *h* classes. Different classes use different values of regularizers because the reliability of training data and web resources of different classes might be different. In K-fold cross validation, each of the *N* training data has been used once as a validation instance. When used as a validation, the decision scores of image i are denoted by $p=[p_1, p_2, \ldots, p_h]^T$ and $t=[t_1, t_2, \ldots, t_h]^T$ respectively.

Given the class label vector $y_i \epsilon \mathcal{R}^h$ of the image i (i.e. a column vector with value 1 at the c^{th} position and 0 at other positions, and *c* is the class label of image i), we intend to make $y_i$be close to its target y_i. Since this optimization is not a straightforward linear optimization problem, it cannot be solved using linear algorithms such as least squares method. Notably, the input data to the optimization function is *h* pairs decision scores obtained from the bimodal classifiers. Thus, we conduct *h* separate optimizations, where the regularizers are derived separately.

Experiments and Results:

In this section, we assess our image data and web resources fusion model for image classification using two benchmark datasets: UIUC-Sport dataset, LabelMe8 dataset. Comparisons of our method with other state-of-the-art methods are conducted.

(1) Results of UIUC-Sport dataset:

The UIUC-Sport dataset contains 8 sports event categories: rock climbing (194 images), badminton (200 images), bocce (137 images), croquet (236 images), polo (182 images), rowing (250 images), sailing (190 images), and snowboarding (190 images). The image number in each class ranges from 137 to 250, and there are 1579 images in total. Some example images in the dataset can be found in Fig. 3. Note that we use the same experimental settings: randomly select 70 images from each class for training and 60 images for testing. Here, we report the mean and standard deviation of the classification accuracy over 30 training/testing random splits.

In Table I, we compare the classification results of the state-of-the-art feature descriptors upon the UIUC-Sport dataset when they are used alone or combined with web resources under the multimodal fusion framework. Obviously, significant improvements are achieved in all the 4 feature extraction methods. Without using any pre-trained deep networks, the benchmark result of the UIUC-Sport dataset was achieved by our fusion method with an overall accuracy of 88.19 ± 1.25%. Our new method using PHOW descriptors produces even better results as shown in Table I.

TABLE I
CLASSIFICATION ACCURACY (%) OF THE STATE-OF-THE-ART FEATURE EXTRACTORS (AVERAGE OVER 30 TRIALS) ON UIUC-SPORTS DATASET. THE BOLDFACED NUMBERS DENOTE THE PERFORMANCE WITH THE OUR TEXT-BASED REPRESENTATION.

Algorithm	Visual feature	Method in [1]	New method
GIST	64.15 ± 1.95	77.44 ± 1.94	$\mathbf{85.76 \pm 1.21}$
ScSPM	80.28 ± 0.93	85.91 ± 0.92	$\mathbf{91 \pm 1.08}$
OB	77.87 ± 0.91	86.98 ± 1.01	$\mathbf{88.21 \pm 1.25}$
PHOW	83.95 ± 1.11	88.19 ± 1.25	$\mathbf{91.11 \pm 1.04}$

Table II shows the classification performance for each class when PHOW descriptors are used without fusion, fusion with the method in [1], and fusion with the newly proposed method. Although the fusion method in [1] has already achieved substantial improvements in overall, the accuracy of class *badminton* slightly drops. However, the new method presented in this report achieves considerable improvements in all classes as shown in the Table II. Among these categories, the result of class *bocce* is improved by 13%, which is a much higher improvement than all previous reported works on UIUC-Sport dataset.

TABLE II
PER-CLASS ACCURACY (%) COMPARISON OF THE STATE-OF-THE-ART FEATURE EXTRACTORS ON UIUC-SPORTS.(AVERAGE OVER 30 TRIALS).

	rockc.	badmi.	bocce	croqu.	polo	rowin.	sailin.	snowb.	mean Acc
PHOW only	93.1	91.6	63.5	76.1	84.7	88.2	90.7	81.3	83.9
Method in [1] (w/ PHOW)	97.2	87.9	66.4	76.8	90.7	96	98.3	89.9	88.2
New method (w/ PHOW)	98.3	95	76.1	83.3	90.6	95.5	99.4	90.1	91.1

(2) Results of LabelMe8 dataset:

The LabeMe8 dataset contains 8 natural and cultural landscape categories: coast (360 images), forest (328 images), highway (260 images), insidecity (308 images), mountain (374 images), opencountry (410 images), street (292 images), and tallbuilding (356 images). The image number in each class varies from 260 to 410, and there are 2688 images in total. Fig. 4 shows some example images in the dataset. We randomly select 100 images from each class for training and another 100 images for testing. Again, the results reported here are the averaged accuracy over 30 training/test splits.

The classification results of the state-of-the-art feature descriptors with or without the multimodal fusion are given in Table IV. The new method with ScSPM descriptors achieves the best result of 93.8 ±0.72%, while the method in [1] slightly improves the overall accuracy. Tabel III shows the per-class performance comparison using ScSPM image descriptors. In contrast to the performance of the method in [1] (no improvements in 4 classes), the new method developed in this report achieves substantial improvements in all the scene classes.

TABLE III
PER-CLASS ACCURACY (%) COMPARISON OF THE STATE-OF-THE-ART FEATURE EXTRACTORS ON LABELME8.(AVERAGE OVER 30 TRIALS).

	coast	forest	highw.	insidc.	mount.	openc.	street	tallb.	mean Acc
ScSPM only	84.2	96.4	90.2	91.5	90.5	72.4	91.7	92.4	88.3
Method in [1] (w/ ScSPM)	85.1	95.1	99	85.8	96	68.5	95.1	91.1	89.1
New method (w/ ScSPM)	96.1	98.3	95.5	95.4	94.4	79.4	95.7	94.8	93.8

By comparing the experimental results of two datasets vertically, we find that the performance of the method in [1] is more sensitive to data, while the new method produces more robust performance in the general situations of scene classification. This reveals that the semantic information underlying the text is indeed meaningful for the understanding of the image contents.

TABLE IV
CLASSIFICATION ACCURACY (%) OF THE STATE-OF-THE-ART FEATURE EXTRACTORS (AVERAGE OVER 30 TRIALS) ON LABELME8 DATASET. THE BOLDFACED NUMBERS DENOTE THE PERFORMANCE WITH THE OUR TEXT-BASED REPRESENTATION.

Algorithm	Visual feature	Method in [1]	New method
GIST	79.36 ± 1.09	82.09 ± 1.18	$\mathbf{91.45 \pm 1.02}$
ScSPM	88.31 ± 1.21	89.12 ± 1.12	$\mathbf{93.8 \pm 0.72}$
OB	84.97 ± 1.01	85.46 ± 0.93	$\mathbf{93.63 \pm 0.88}$
PHOW	87.58 ± 1.02	88.14 ± 1.01	$\mathbf{92.04 \pm 0.88}$

(3) Discussion:

In this report, we have proposed an adaptive multimodal fusion framework that uses both training data and web resources for scene classification. Experimental results on the benchmark datasets show that the proposed text-aided scene classification framework could significantly improve classification performance. Experimental results also show that the adaptive multimodal fusion mechanism can effectively individualize reliability-dependent weighting modulation for every new observation.

The web resources obtained from online search is not limited to any particular scene classification task. This characteristic creates a prospect of recognizing images from unseen new classes, which is needed in many visual recognition applications. Recognition of unseen new scene classes using the proposed framework is under exploration, and results will be reported in our future publications. In addition, some advanced adaptive fusion approaches such as context-aware fusion will be explored so that context-adaptive weights could be assigned to different classifiers to boost fusion performance.

1. List of Journal Publications and Significant Collaborations that resulted from your AOARD supported project:

[1] Xuefeng Yang, Kezhi Mao. (2016). Task Independent Fine Tuning for Word Embeddings. IEEE Transactions on Audio Speech and Language Processing, 2016, Accepted.

[2] J.Wang, J.Xie, Rui Zhao, Kezhi Mao, and L.Zhang. (2016). A New Probabilistic Kernel Factor Analysis for Multisensory Data Fusion: Application to Toll Condition Monitoring. IEEE Transactions on Instrumentation and Measurement, 65(11), 2527-2537.

[3] Rui Zhao and Kezhi Mao. (2016). Topic-Aware Deep Compositional Models for Sentence Classification. IEEE Transactions on Audio Speech and Language Processing, 2016, Accepted.

[4] Rui Zhao and K.Z. Mao, Cyberbullying Detection Based on Semantic-Enhanced Marginalized Denoising Auto-Encoder, IEEE Transactions on Affective Computing, accepted, 2016.

[5] Xuefeng Yang and K.Z. Mao, Learning Multi-Prototype Word Embedding from Single-Prototype Word Embedding with Integrated Knowledge, Expert Systems with Applications, accepted, 2016.

2. Invited talks that resulted from your AOARD supported project:

[1] Zhibo Xiao, Tharini Nayanika de Silva, Chen Wei, Kezhi Mao, Evolving Knowledge Extraction from Online Resources, "ICMLC 2017: International Conference on Machine Learning and Cybernetics, London, July 2017, accepted for presentation.

[2] Tharini Nayanika de Silva, Xiao Zhibo, Kezhi Mao, Causal Relation Identification Using Convolutional Neural Networks and Knowledge Based Features, ICMLC 2017: International Conference on Machine Learning and Cybernetics, London, July 2017, invited for presentation.

[3] Dongzhe Wang, K.Z. Mao, Gee-Wah Ng, Tien Pham, Adaptive Multimodal Fusion of Training Data and Web Resources, Proceedings of 19th International Conference on Information Fusion, Heidelberg, Germany, July 2016.

[4] Zhibo Xiao, Tharini Nayanika de Silva, Chen Wei, Kezhi Mao, Gee-Wah Ng Constructing Bayesian Networks by Harvesting Knowledge from Online Resources, 19th International Conference on Information Fusion, Heidelberg, Germany, July 2016.

[5] Dongzhe Wang, K.Z. Mao and Gee-Wah Ng, Improving scene classification by fusion of training data and web resources, Proceedings of 18th International Conference on Information Fusion, Washington D.C., USA, July 2015.

[6] Xuefeng Yang, Rui Zhao and K.Z. Mao, Distributional Sentence Representation by Expert Knowledge for Causal Relation Identification, Proceedings of 10th International Conference on Information, Communications and Signal Processing, Singapore, Dec. 2015.

[7] Xuefeng Yang, Rui Zhao and K.Z. Mao, Regularized Training of Compositional Distributional Semantic Models, Proceedings of 10th International Conference on Information, Communications and Signal Processing, Singapore, Dec. 2015.

[8] Dongzhe Wang, Rui Zhao and K.Z. Mao, Partially Connected ELM for Fast and Effective Scene Classification, Proceedings of 2015 International Conference on Extreme Learning Machine, Hangzhou, China, Dec. 2015.

[9] Rui Zhao, Anna Zhou and K. Z. Mao, 17th International Conference on Distributed Computing and Networking, 1st International Workshop on Understanding Situations Through Multimodal Sensing, Singapore, Jan. 2016.

[10] K. Wu, Wenyin Tang, K.Z. Mao, G.-W. Ng, Semantic-level fusion of heterogeneous sensor network and other sources based on Bayesian network Semantic-level fusion of heterogeneous sensor network and other sources based on Bayesian network, Proceedings of 17th International Conference on Information Fusion, page 1-7, Spain, July, 2014.

[11] Rui Zhao, Kezhi Mao, Supervised Adaptive-transfer PLSA for Cross-Domain Text Classification, 14th IEEE International Conference on Data Mining Workshop in Domain-driven Data Mining, China, December, 2014.

3. Grants received that are related to your AOARD supported project:

No.	Funding Source	Quantum ($)	Role	Grant Period	Project Title
1.	Ministry of Defence	$755,785	PI	2016-2019	Automated Situation Awareness Based on Text Analytics of News
2.	Ministry of Defence and NTU	$80,000	PI	2016-2017	Exploration of Multisensory Integration for Adaptive Activity Recognition
3.	Ministry of Defence	$523,372	PI	2013-2016	Automatic Knowledge Extraction from Unstructured Open Sources
4.	US Department of Air Force, AOARD	$195,925	PI	2013-2016	Fusion and Sense Making of Heterogeneous Sensor Network and Other Sources
5.	DSO National Laboratories	$35,000	PI	2014-2015	Cognitive Fusion and Sense Making of Heterogeneous Sensor Network and Other Sources