An Approach to Increase Productivity and Safety in Garments Using Deep Learning and Image Recognition

Shafkat Waheed
Research Assistant
Electrical and Computer Engineering
North South University
shafkat.waheed@gmail.com

B. M. Raihanul Haque
Research Assistant
Electrical and Computer Engineering
North South University
raihanul.haque@northsouth.edu

Dr. Mohammad Ashrafuzzaman Khan
Assistant Professor
Electrical and Computer Engineering
North South University
mohammad.khan02@northsouth.edu

***Abstract*—This paper investigates the application of Image captioning in garments sector to increase production and provide safety from workload imbalance.Though Bangladesh is a second largest garments manufacturing country, it often fails to meet the production goals in due time.Moreover It is unable to meet the safety standards of the industry losing contract from International buyers. In this paper we outlined the application of image caption to detect complex work using non invasive CCTV camera.We also tried to derive time for such work using graph matching.In this paper we showed how time derived from Deep learning neural network and graph matching can be used to increase efficiency and reduce safety issues due to fatigue.Besides this paper represents a mitigation way and a proposed model in production line to tackle problems related to safety and production that will help the top management.**

1. Introduction

Providing safety and security for labours has been a challenging issue all in recent times.The usage and manipulation of heavy machinery in an monotonous setting if kept unsupervised create room for accidents and loss for the company and the labours.The progress of deep learning and image recognition gives us a chance to overcome this challenge without human intervention.

Manufacturing industry is no stranger to Iot devices and A.I[?].Germany was the first to understand the potential of optimizing manufacturing process using IoT devices and Artificial Intelligence[?].They were able to change the whole scene of manufacturing with the integration of small devices in everyday production.Sensors like accelerometer, gyroscope, heat detector, light detector and vibration detector increased the dimension of standard information one could garner or gather. Information of such volume crafted the way for machine learning and A.I to effectively optimize the work flow, industrial production and efficiency.

2. Related Works

Bangladesh is the second largest exporter of garments.More than 80 percent of the export in this country comes from garments, which contributes to 15 percent of the GDP.Though the maturity of this industry is very high in Bangladesh.Over the years techniques have been developed to improve the efficiency and safety of production using tested industrial techniques.Kader [1] analyzed the factors effecting lead time in ready made garments.The researchers discovered that Bangladesh even being one of the second largest exporters of garments suffered greatly on its lead time.The researchers proposed a three stage step strategy that involve management of production in due time through supervision.M.M khatun [2] tried to fine tune the production process by diving the work and time management to unit components.She suggested time management process should be divided into worker capacity which involves process wise calculation of capacity,M.M khatun [3] further analyzed her procedure and tried to find relation to time and motion in productivity of garments industry.S.S Jadhav and G.SSharma [4] added on the thought of time based productivity.They researched on supply of garments piece and suggested that proper timing of supplied components can improve productivity in garments industry.Productivity is important for any garments to profit and produce high quality products.But these productivity can also be hampered with sudden accidents occurring in the production process.These incidents not only hamper the production rate but also risk workers with injuries that are detrimental to health and life.J C Hiba [5] wrote a manual on safety issues and guideline to follow in garments.The guideline outlines the problems faced in the garments industry and how to tackle them.Most of the issues could be handle through management but guidelines of fatigue and rest are mostly ignored or not properly handled.M. Ahmed and G.Kibria [6] worked on a process called 6S study.His research showed how 6S improved productivity by 27 percent and multi factor productivity by 13 percent in the work process.The safety issues suggested in his research puts liability on unorganized workforce and long hours of work in these

environments.

These issues of safety and productivity is directly likened to time and its proper utilization in governance of the factory.Technology such as machine learning and deep learning allows us to optimize the use of time and push the boundaries for these old traditional methods of optimization.As we are moving towards the Fourth industrial revolution technologies such as Iot,big data and deep learning are becoming more and more necessary [7].One of the ways of optimizing time is knowing what work an worker is performing.Taha [8] outlined how data from phone can be used to predict human activity with the help of traditional machine learning.They collected data from the accelerometer of phone.The collected data was used to map the work a person is performing.Batilak [9] used a different approach,they used RGB camera and Kinet to detect human activity for surveillance.The model took frames of images with kinet and used them to detect human activity.They applied SVM to classify and detect the work.But it had limitation as it could only detect activity of short time period.Lu [10] used a different approach,They instead of only suing one medium of data used two sources of data.They used CapsNet and Lstm to extract features from tow sets of data originating from camera and sensors to accurately classify nine activity of short time span.Laput [11] instead of using any image they solely used wearable to determine hand gestures, input modalities, motor powered object detection and so on.The combined system is called ViBand.AJ Piergiovanni [12] followed a different strategy instead of using supervised methods, they used unsupervised learning to extract features from raw data collected suing sensors.Later they used RBF-SVM to to classify activity.All these techniques of activity of detection are great from detecting activity but these detection requires wearable devices or costly device like kinet to classify activity.These types of classification limits us to segment the work process into categories which hinders the ability to generalize the detection process.AJ Piergiovanni and Co. [12] discussed an approach to classify human activity using temporal attention filters. Any high-level activity can be sub-divided into multiple small events also known as sub-events or temporal events and they can vary in terms of duration. From a video feed, with the help of temporal filters coincided with segment based CNN and recurrent LSTM, the system learn these small events which corresponds to certain activity.This approach removed the need of pre-emptive classification of the work rather the network defines the work based on image.G Li, L Zhu, P Liu and Y Yang [13] used attention mechanism to denote images with caption which describes the image in context to its actual representation.This researchers showed how images can be represented in language through the use of deep learning without the mapping of features to classification.

We used deep learning to find a less invasive approach of detection of an complex work as well as detect the work to produce a time frame through which we can apply the already established traditional approach of increasing productivity and safety for the industry and workers.

3. Industry Visit

We visited a garments manufacturing plant to get a better grasp of the issues faced by the labour.We saw various machinery being manipulated by worker for long hours.The factory was huge and had a lot of area to cover in the respect of supervision and monitoring.Therefore,in this huge area it becomes hassle for top management to properly care for workers to check for fatigue or manage delay in production as it happens.A good monitoring system that can inform supervisors quickly of the delay in production occurring in real time, worker working overtime or breaking standard protocol is bound to boost the work environment.

4. Proposed Methods

To carry out the task few machine learning models have been considered. Each model has their own role to play and also generate results collectively. From end to end, two neural networks i.e. Convolutional Neural Network (CNN) and Recurrent Neural Network (RNN) have been used along with a Fourier transformation strategy called wavelet analysis. eSense data and images will be analyzed by these suggested methodology to produce comprehensive and meaningful results.

4.1. CNN

One of the benefits of using this network is, unlike other algorithms, less pre-processing is required which makes it much more suitable for image analysis. The other convenient factor is it can capture spatial and temporal dependencies while reducing number of parameters and reusing weights to understand a sophisticated image. The dimension of the input image is denoted as such, $height \times breadth \times channels(RGB)$ and after performing convolution operation with a kernel or a filter a new matrix having convoluted features is generated. Whether there is one channel or are multiple channels the kernel strides in such a way which creates squashed one-depth channel convoluted feature output. The first layer extracts low level features and the following layers extracts high level features which creates a network to understand a image like a human would do. However, the dimensional reduction process is carried out by valid padding and in order to preserve the same dimension or increase it same padding is incorporated. Pooling is of two type i.e. max and average and this technique is brought to extract dominant feature and decrements the spatial size which allows to use low computational power. Max pooling is preferable since it is noise suppressant. After going through all these processes, the obtained values are flattened into a column vector which is then fed to a conventional feed-forward neural network along with back-propagation techniques.

4.2. RNN

Both CNN and RNN have fundamental similarities which is sharing parameters. RNN has the ability to gen-

Figure 1: Factory Visit

erate future information based on its past. A general NN remembers things during training and while RNN does the same, additionally, it remembers stuffs from previous inputs during producing outputs. Also, unlike NN, RNN can tackle unlimited number of inputs (not fixed initially) and these input vectors are manipulated by the weights of the inputs and hidden state vectors. Thus, this can give rise to one or more output vectors. Since, no fixed input is fed into this model there cannot be any fixed weight for individual input. Thus weights are being shared by each input and to maintain versatility and depth hidden state vectors come into action creating link between two inputs. This parameter sharing strategy makes it different than conventional NN. Furthermore, to have multi-level abstraction and representation any of the four following methods can be tried; (a) have more hidden states, (b) have more non-linear hidden layers and lay them between input and hidden state, (c) have more depth within hidden states and (d) have more depth in between hidden states and output layer. These techniques can also be found in Bidirectional RNN, Recursive neural netwrok, Encoder Decoder Sequence to Sequence RNN and last but not least in LSTM with slight variation.

4.3. Transfer Learning

Transfer learning is a concept in deep learning where a model can leverage knowledge i.e. features, weights to another model. Instead of building models from the scratch for similar but new tasks, pre-trained models can be used. For instance, if a model is built to tackle NLP (Natural Language Processing) related task in English language, the same model might be used for German language. One such model is VGG16 which, in its core, has convolution neural net (CNN) architecture. It has convolution layers of 3x3 filter with a stride 1 and padding and maxpool layer of 2x2 filter of stride 2. It is used to handle computer vision based problems such as image classification, image captioning, feature extraction etc.

4.4. Show,Attend and Tell

This methodology is proposed by Kelvin Xu et. al. which is able to automatically detect distinct objects and generate caption from an image. This model contains a CNN-LSTM network. This types of captioning not only require to understand the objects in an image but also relation between them. previously, fully connected layer system has been used to detect features. But, in this paper a lower level CNN is used to extract features. One the other hand, the LSTM has been used and trained in sequence to sequence manner to generate words. Show, Attend and Tell uses not only CNN but also soft and hard attention techniques. This paper uses VGG architecture which is a pre-trained model to generate feature maps. This later gets converted into vectors. This features are then sent to LSTM with attention model. Attention is a feature of the system that causes the system to find and extract most obvious features of an image. However, LSTM with attention model has multiple gates. The inputs are modified before going through the next gate. The generated vectors are called context vectors. The formula for generating vectors is: $e_{ti} = f_{att}(a_i, h_{t-1})$, $\alpha_{ti} = \frac{exp(e_{ti})}{\sum_{K=1}^{L} exp(e_{tk})}$ and $\hat{z} = \phi(a_i, \alpha_i)$. Predicting next word in the caption is done by this formula, $p(y_t|a, y_1^{t-1}) \propto exp(L_0(Ey_{t-1} + L_h h_t + L_z \hat{z}_t))$. Going back to attention, hard attention is complex technique where the location with feature, is sampled from a multi-nomial distribution. As opposed to this soft attention is much more trivial. Finally, the model is evaluated using BLEU(Bilingual Evaluation

Understudy) score. Initially, BLEU has been developed to measure machine translation. But, in the paper it is used to compare between predicted captions and reference captions. The result excelled the result obtained from previous model.

5. Data

We used three data-set to build the model for providing safety and provide efficiency in garment.Mainly, The first data-set was a video of poaching an egg.We basically, recorded it to build a test run for the neural network.It involved a simple action of poaching and egg in a stove.We used it to the pre-formulate the idea of the model. Second video was collected from YouTube,It showed a person performing simple task in an industry. It involved picking up object and putting in the designated location. The third video was collected from YouTube which showed the whole procedure of garments manufacturing [14].

6. Methodology - Pre-Model Formulation

Videos are basically images changed frame by frame [15]. We fragmented the video stream into video frames.Figure 2 represents the generalization of the video frame procedure.

Figure 2: Video feed to video frames

The objective here was to generate a description of a complex work that was retrieved from a video feed. To accomplish that, initially we captured a video of a poached egg to determine the capability of the network. The process of poaching an egg consisted of eight steps where two particular steps are identical to another two. The duration of each step was 10 to 12 seconds long so that we could avoid over-fitting. At first, image were extracted from these video feeds. Then we used transfer learning to extract features from images using VGG19 [16].After extracting the features from images we captioned the images based on the work they represented.

6.1. Image extraction

To [1] extract the images,we used one of the most popular library known as OpenCv [17].As such, we generated frames from video using this library. Total number of 650 frames were extracted from the video. Figure 3 represents one of the frames that were extracted.

6.2. Feature Extraction

We passed each of the images through a CNN to extract features that would represent important part of the images.In addition,we visualized the feature map to fully understand what part of the images CNN was learning.Moreover,in figure 4 we showed the feature maps and what features were collected from the image.

Figure 3: A frame of the complex work showing egg is being hold by a person

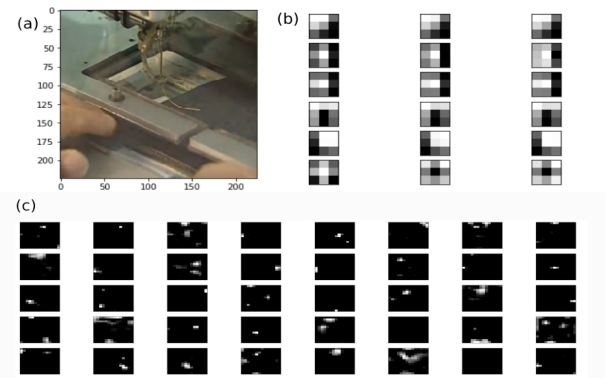

Figure 4: Feature map and feature extraction

Figure 4(a) we showed The picture of sowing and figure 4(b) represents the feature maps.The feature maps extracted information like edges,shape and corner represented in the image.In figure 3(c) we showed the parts of the images that was used as important features.

6.3. Image captioning

At the beginning, unique ids were created for each image and against each id two captions were assigned. Since, a bunch of images represents a particular phase captions generated for those images are same. Now, to extract the features from images VGG model is called upon which is a pre-trained model. VGG causes faster computation and less memory consumption. The focus here revolves around

returning the dictionary which contains image features of internal layers and save it to a file.

The captions that were generated earlier were loaded from a file and at the same time unique frame identifier is returned based on the description. To alleviate the difficulty working with the description, they were tokenized and cleaned. Tokeniztio [18] is a procedure where description turns into individual words or vocabulary and cleaning process involved making the description in lowercase and remove redundant words, numbers, punctuation marks and articles. This newly created dictionary is then saved to a new file to be loaded later on.

At the beginning of our training, photo features were loaded from the file that was previously created. Along with that, description was loaded ase well. In order to map description to unique integer value tokenization was performed on the training set before feeding it to the model. Two string, 'startseq' and 'endseq', were used to mark the start and end of an caption. The sequence of words were generate based on the parameters i.e. highest length of the sequence, tokenizer and both descriptions (image and text).

As for the model, Long Short-Term Memory Recurrent Neural Network (LSTM-RNN) with attention were implemented. The text, which was encoded as integers, was fed to one part of the model and the image is fed to another part. This will create a probabilistic distribution of the caption to be matched with an image [13].

A test dataset was formed containing photos. Then, the model was called recursively to generate captions against the test dataset. Then, a mapping between the caption and image was observed to see if the model can successfully generate the description. Figure 6 and Figure 7 illustrates how well captions were generated for new dataset.

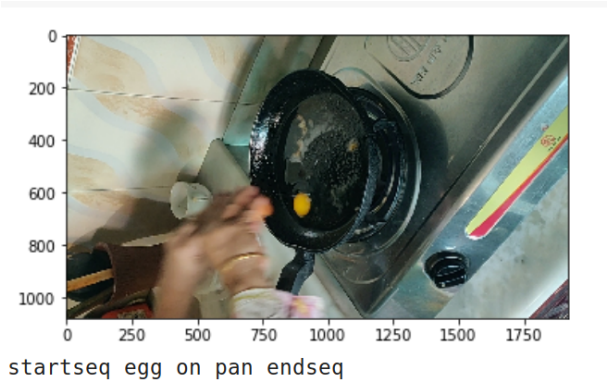

Figure 5: Successful prediction of an image from test dataset

6.4. Graph generation

Before generating the graph, the descriptions were split into noun and preposition.We made a list of preposition and a dictionary containing pairs of nouns to be connected by any of the preposition from the list. Each node of the graph represents noun and each directed edges the preposition. The number associated with preposition indicates the position of

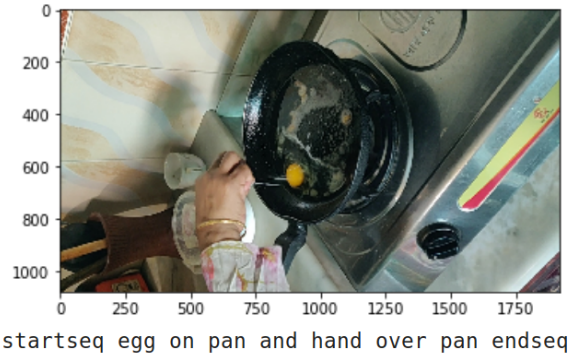

Figure 6: Successful prediction of another image from test dataset

the word in the sentence. For all the similar captions there is only graph mapped to it, illustrating a certain steps. Figure 8 demonstrate the outlook of the graph for a certain caption.

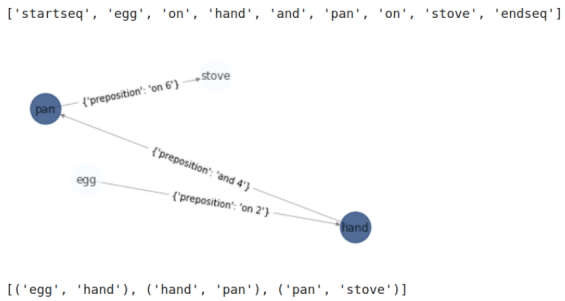

Figure 7: Directed graph of a caption

we wanted to match the generation of the graph in a continues loop to determine the work process and how long it was performing.Even thought, the model tends to give near similar result in some instance it was necessary for us to generate a graph and match the two instances to determine a similarity of the caption per frame.

7. Methodology - Model Formulation

In the pre-model formulation we tried to find the possibility and applicability of the image captioning technique.We were successfully able to achieve good decisive result. We trained and applied our network on a small video from YouTube to test the comprehensiveness of deep learning model.We converted the video feed into video frames.Furthermore, we divided the frames into a train, validation and test set.After training we decided to test our model to see the caption and determine its capability.

Our main goal in this stage was to observe, if the work of putting object in different place could be captioned.As such,we were successfully able to caption the states of activity in the video.In Figure 9 we showed the caption generated and it was successfully able to caption the activities the person has performed.

7.1. Derivation of the Model

Initial trials produced good prediction on the complex work.As such,we derived a model that generated caption to determine the activity.Which involves, how a worker is performing and how it integrates itself with the working environment to answer question in relevance with industry established methods.In addition,Time delay and Work time helped us to flag key parts of the work process to automate the garments evaluation process.In figure 9 we can see a full picture of the model for automation.

We found the Work Time by matching of graphs show in figure 9 generated by each frame.After knowing, the time of the work we determined or calculated the time delay for a particular activity.

$$TimeDelay = OptimumTime - WorkTime \quad (1)$$

In equation(1) Time delay was the lag in performing an activity while Optimum time is the known value for performing a work on time.In addition to that,Work time is the time an worker was actually involved in doing the activity.Work Time is a good indicator of determining whether a person is doing overtime or not.As such,doing overtime creates fatigue in workers, making them sluggish and less responsive to the environment.Moreover,Workers handle complex and dangerous machines to perform their tasks, which if not properly managed in full awareness would create serious life threatening accidents.

8. Methodology - Garments Industry

Garments workers perform various segmented work to produce a garment. Nevertheless,we focused on the activities which mainly faced accidents and delay due to poor supervision [19].These are activities like sewing a cloth, cutting a material for making garments components and dyeing clothes for giving the their vibrant color. We followed all the steps used in previous phases,The distribution of the data-set, according to the caption.

8.1. Sow

Sewing involves piecing the components together.Generally,components of clothes are sowed to give its appearance and value.We captioned the work using our model and calculated the time of his work, figure-11 represents the overall detection of sewing video collected from YouTube.We passed the video through our model and it was able to caption the work that was being preformed.In addition,We matched the similarity to calculate the Work Time.

On figure 11,we can see the features being used to generate the text.Furthermore the graph is generated using the caption.We observed that same graphs are generated as long as person is performing the activity.We calculated the time by multiplying the video frame number with its FPS.We knew the optimum time for this activity, we calculated the lag in sewing using the equation 1.Time delay is a good indicator of much are we lagging behind in production.

8.2. Cut

Cutting is very much necessary to make components necessary for sewing.We evaluated the activity using our model. Figure 12 represents the overall detection of cutting.The was video collected from YouTube and we transitioned the video through our model and was able to caption the work that was being preformed.likewise,We matched the similarity of the graph produced and calculate the Work Time.

On figure 12, we showed the video frames and caption generated through the model.Like before,we were able to pro-grammatically determine how long he was performing the activity by multiplying the graphs generated with FPS and determined the delay in his activity using equation 1.

8.3. Dye

Dyeing is necessary to give clothes its vibrant color.It involves spraying color onto clothes to make it look unique and appealing.Similarly, like cutting and sewing we detected this work using our model.In the same manner,We calculated the work time figure-13 represents the overall detection of cutting.The was video collected from YouTube. As,we passed the video through our model and we were able to caption the work that was being preformed just like sewing and cutting.

We evaluated if a person is suffering from fatigue or not by observing the work time. [20]In modern day,we can avoid fatigue in work place through proper monitoring and scheduling.Likewise,Its is the job of the supervisor to monitor this workload to maintain a productive work force.Therefore,Work time computed using the image recognition technology would act as a supplement for the solution to this problem of management.

9. Result

In the methodology pre-model section, we have mentioned an event or action (poached egg making) and how we extracted the data, generated captions and produce graph of the corresponding image.On the Model formulation section we used a YouTube video to see the ability of the neural network. The video demonstrates a man, organizing tools (wrench, hook etc.) as part of maintenance work. Using the same process, we have generated graphs of the corresponding images as shown in Figure 6. One thing to note here is that in some of the images i.e. c_1, e_1 and i_1 noise is present (graphical tick mark). This is because we used a video that is not recorded by us. In real life cases this type of noises can easily be avoided.

Figure 6 represents that, in all the 9 cases the graphs have been generated successfully. Usually, in real life scenario

many incident takes place withing a certain time frame and it is possible to generate captions for all of them. However, this could be redundant and we only generated graph captions for those which are relevant to the event. For instance, if the man putting a hook on the bar then the captions will be like hook, on ,bar with arrows in between.

Unlike our poached egg example, the nodes of opposite ends do not represent noun only but any other types of word. For example, in b_2 the caption is perfectly, on ,bar. This shows the efficiency of graph generation in fashion which can describe an event properly. Likewise, the middle word can be a verb instead of a preposition as seen in h_2.

But, for both the case we used VGG19 to extract features from images but in the third and final methodology, we used Resnet,VGG19, and Inceptonv3 to determine extract images and among them InceptionV3 showed the most promising results.The following table -1 shows the results of caption generation.We used Blue Score to evaluate the CaptionNet.

Table 1: Results

Network Evaluation			
CNN Name	Blue Score 1	Blue Score 2	Blue Score 3
VGG19	0.360800	0.349552	0.187343
Resnet	0.375260	0.360864	0.201157
InceptionV3	0.386725	0.379816	0.215263

From the table 1 its is evident that Resnet offers the best results compared to the other CNN models.The reason being the deep nature of its architecture [21].

10. Conclusion

The proposed method is a detection system that is able to provide safety and security to factory worker as well as improve production in the factory.Moreover,the method highlights how it is possible to use deep learning in industry to work as safety net to handle imposing accidents caused due to overwork or fatigue and production hindrance that could arise for not following proper rules and regulation of industry.Even so,The model could improve further if the data set was more rich and diverse.Image recognition is cheap and less invasive, giving it more room for application in industry 4.0.Specially in sectors where monitoring is extremely necessary.

The method was tasted on real world scenario to show the possible usage of the technique.We plan to improve our technique and further extend its functionality by adding additional source of data collected from various sensors.We hope such addition might increase the reliability of the model to be used in large scale.

Acknowledgments

References

[1] S. Kader and M. M. K. Akter, "Analysis of the factors affecting the lead time for export of readymade apparels from bangladesh; proposals for strategic reduction of lead time," *European Scientific Journal*, vol. 10, no. 33, 2014.

[2] M. M. Khatun, "Application of industrial engineering technique for better productivity in garments production," *International Journal of Science, Environment and Technology, II*, pp. 1361–1369, 2013.

[3] M. M. Khatun, "Effect of time and motion study on productivity in garment sector," *International Journal of Scientific & Engineering Research*, vol. 5, no. 5, pp. 825–833, 2014.

[4] S. Jadhav, G. Sharma, A. Daberao, and S. Gulhane, "Improving productivity of garment industry with time study," *International Journal on Textile Engineering and Processes*, vol. 3, no. 3, pp. 1–6, 2017.

[5] J. C. Hiba, *Improving working conditions and productivity in the garment industry: An action manual*. Intl Labour Organisation, 1998.

[6] M. Ahmed, T. Islam, and G. Kibria, "Study on 6s method and improving working environments in the garments industry," *International Journal of Scientific & Engineering Research*, vol. 9, no. 3, pp. 737–754, 2018.

[7] E. Manavalan and K. Jayakrishna, "A review of internet of things (iot) embedded sustainable supply chain for industry 4.0 requirements," *Computers & Industrial Engineering*, vol. 127, pp. 925–953, 2019.

[8] A. Taha, H. H. Zayed, M. Khalifa, and E.-S. M. El-Horbaty, "Human activity recognition for surveillance applications," in *Proceedings of the 7th International Conference on Information Technology*, pp. 577–586, 2015.

[9] K. V. Bhaltilak, H. Kaur, and C. Khosla, "Human motion analysis with the help of video surveillance: a review," *International Journal of Science, Engineering and Computer Technology*, vol. 4, no. 9, p. 245, 2014.

[10] Y. Lu and S. Velipasalar, "Autonomous human activity classification from ego-vision camera and accelerometer data," *arXiv preprint arXiv:1905.13533*, 2019.

[11] G. Laput, R. Xiao, and C. Harrison, "Viband: High-fidelity bio-acoustic sensing using commodity smartwatch accelerometers," in *Proceedings of the 29th Annual Symposium on User Interface Software and Technology*, pp. 321–333, 2016.

[12] A. Piergiovanni, C. Fan, and M. S. Ryoo, "Learning latent sub-events in activity videos using temporal attention filters," *arXiv preprint arXiv:1605.08140*, 2016.

[13] L. Chen, H. Zhang, J. Xiao, L. Nie, J. Shao, W. Liu, and T.-S. Chua, "Sca-cnn: Spatial and channel-wise attention in convolutional networks for image captioning," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 5659–5667, 2017.

[14] T. Kütüphane, "Garment construction- jean construction," 01 2016.

[15] A. M. Tekalp, *Digital video processing*. Prentice Hall Press, 2015.

[16] Y. Wu, X. Qin, Y. Pan, and C. Yuan, "Convolution neural network based transfer learning for classification of flowers," in *2018 IEEE 3rd International Conference on Signal and Image Processing (ICSIP)*, pp. 562–566, IEEE, 2018.

[17] G. Bradski and A. Kaehler, *Learning OpenCV: Computer vision with the OpenCV library*. " O'Reilly Media, Inc.", 2008.

[18] P. Mcnamee and J. Mayfield, "Character n-gram tokenization for european language text retrieval," *Information retrieval*, vol. 7, no. 1-2, pp. 73–97, 2004.

[19] S. Akhter, A. Salahuddin, M. Iqbal, A. Malek, and N. Jahan, "Health and occupational safety for female workforce of garment industries in bangladesh," *Journal of Mechanical Engineering*, vol. 41, no. 1, pp. 65–70, 2010.

[20] S. E. Lerman, E. Eskin, D. J. Flower, E. C. George, B. Gerson, N. Hartenbaum, S. R. Hursh, M. Moore-Ede, *et al.*, "Fatigue risk management in the workplace," *Journal of Occupational and Environmental Medicine*, vol. 54, no. 2, pp. 231–258, 2012.

[21] A. Géron, *Hands-on machine learning with Scikit-Learn, Keras, and TensorFlow: Concepts, tools, and techniques to build intelligent systems*. O'Reilly Media, 2019.

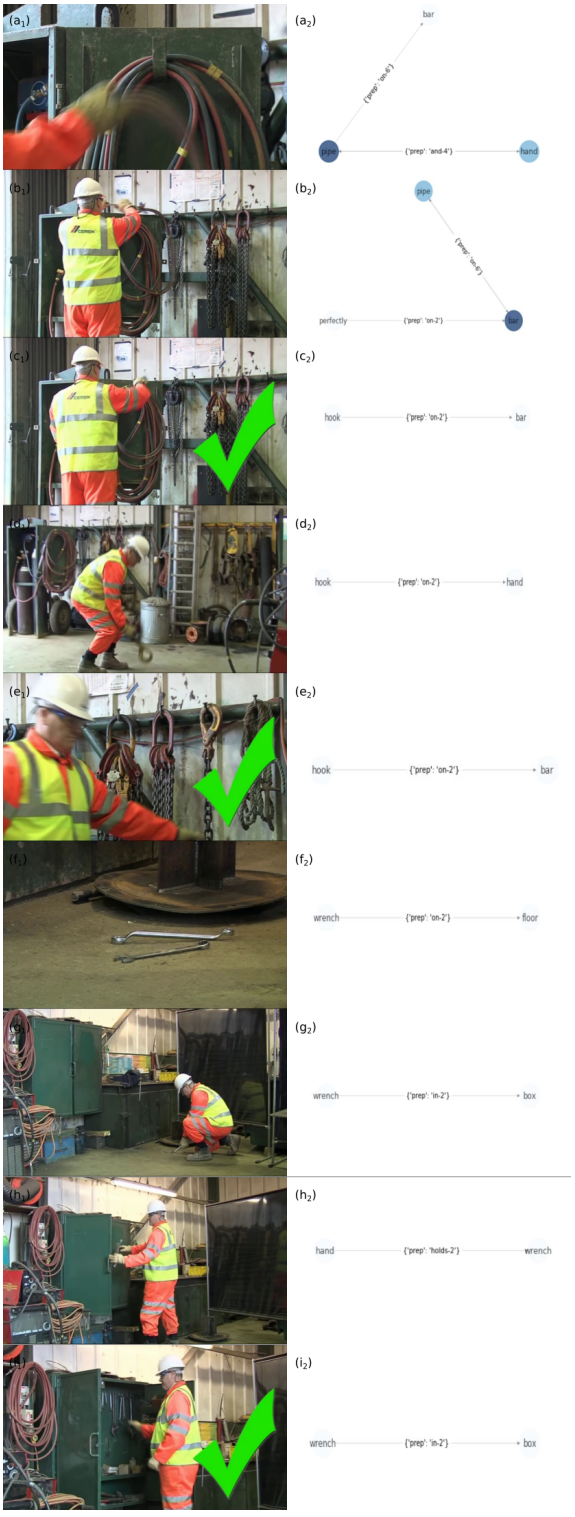

Figure 8: Generated graphs using the frames of a video

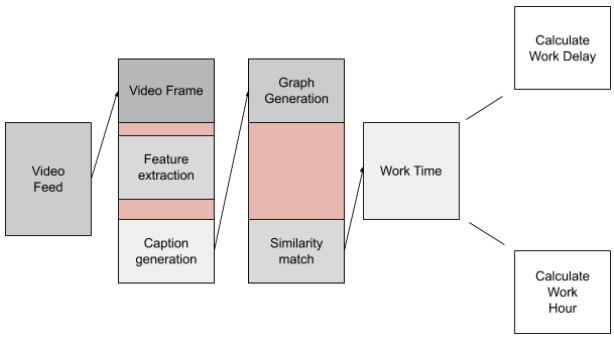

Figure 9: Automation Model

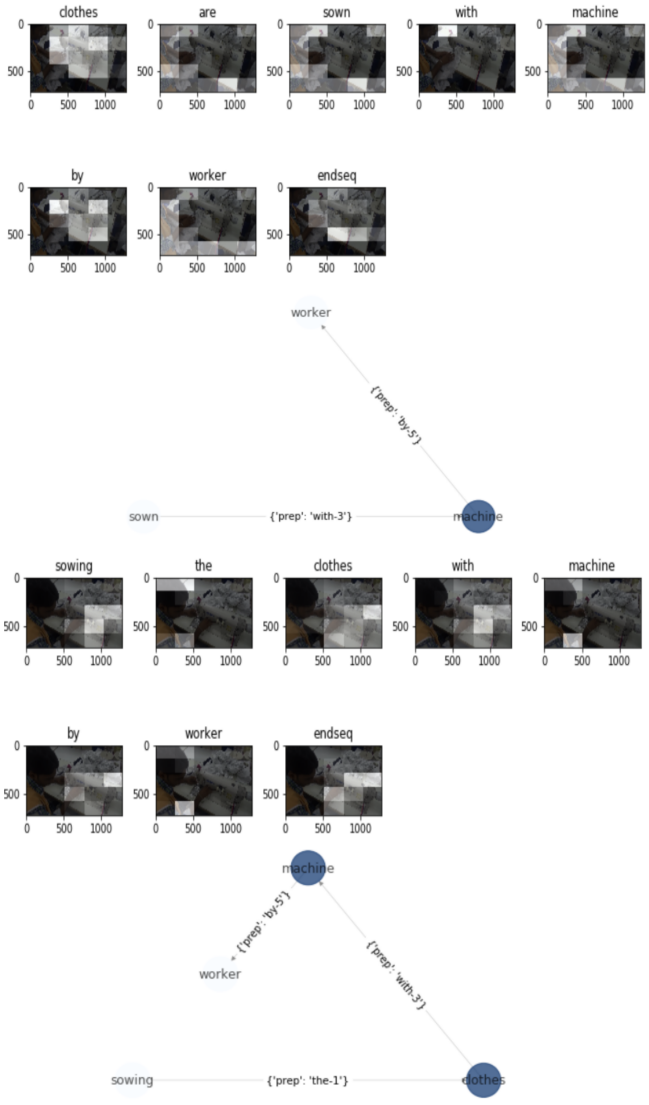

Figure 10: Caption for sewing

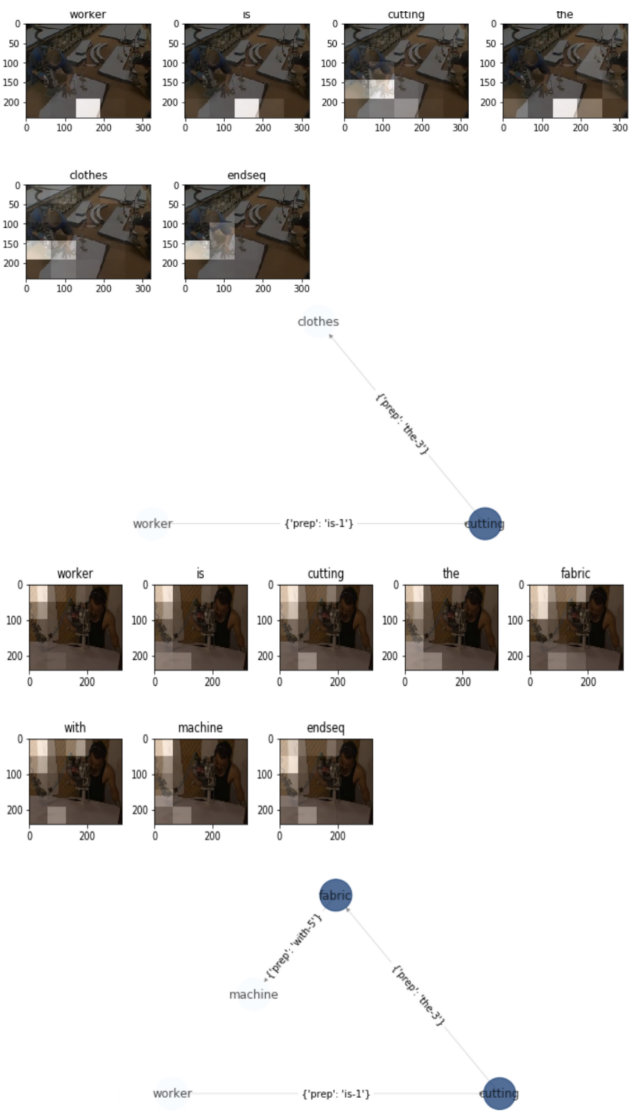

Figure 11: Caption for Cutting

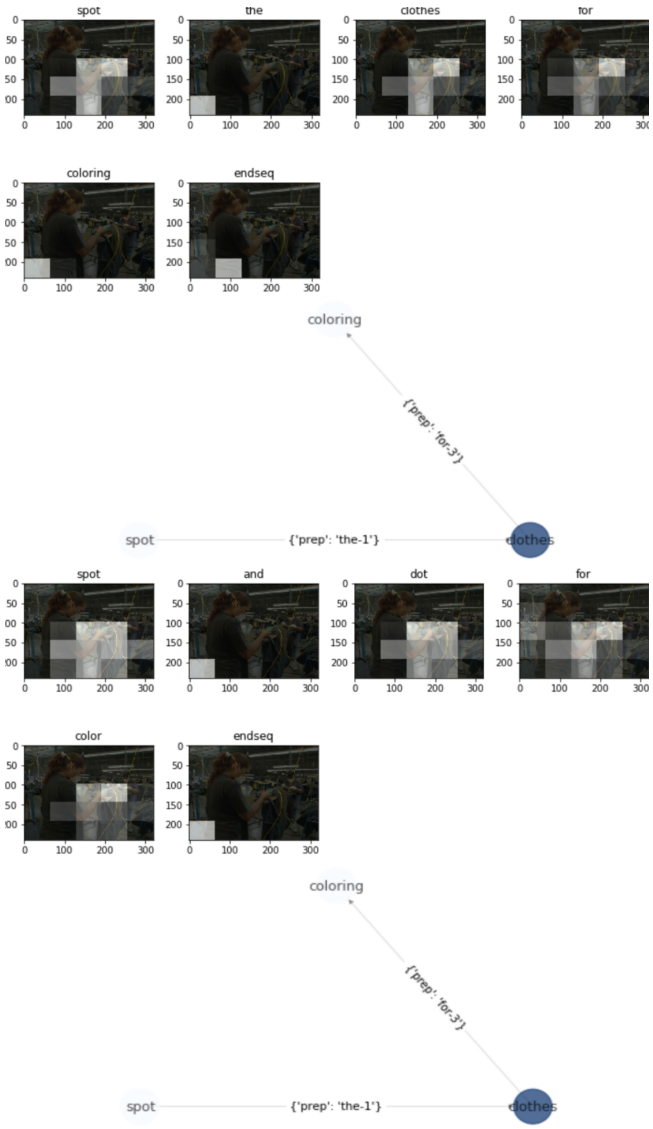

Figure 12: Caption for Dyeing