A Visual Approach to Data Fusion in Sensor Networks

Martin Zsifkovits, Maximilian Moll, Truong Son Pham, and Stefan Wolfgang Pickl

Abstract—Critical Infrastructures such as public transport are essential for every modern society. This makes them even more attractive and vulnerable for terrorist attacks. Recent attacks all over the world demonstrate a strong need of intelligent solutions to detect attackers in an early stage in order to avoid harm to people and the system. Even though privacy rights are important for society, they can also confuse the process of early detection, as many countries limit the use of personalized analysis systems. We therefore introduce a novel approach for data fusion in sensor networks that processes anonymous data from distance based sensors such as explosive detectors. The aim of our ongoing study is not only to detect the presence of an attacker carrying an explosive at a given time, but also to analyze movement over time in an learning environment. For the graphical representation we apply heat maps on a grid overlaying to the given train station.

Index Terms—Distance based detection, Explosive's Detection, Security, Critical Infrastructures.

I. INTRODUCTION

PUBLIC transportation systems are frequent targets to terrorist attacks. The reasons therefore are versatile. In general, critical infrastructures are essential for modern societies. In transportation we further more have high volumes of people directly involved on a daily basis. Thus, a disruption of the system affects an extremely high number of people, especially in urban areas. An attack in such system therefore has the power to cause a high rate of fear and uncertainty within society, which is one of the targets of several terrorist groups. For increasing the security in those systems, two main approaches are possible and applied in reality. Firstly, actively following and chasing groups of potential attackers and arresting individuals before they go into action. Secondly, implementing security measures such as sensors or security staff to passively secure the network. In our research we only consider the second approach. Here, providers of such systems are facing various regulations and laws over different countries that limit the applications of security measures and technologies in order to protect ground rights. This situation strongly demands a solution based on anonymous data within an intelligent system.

We therefore propose a learning system that bases on sensor data over time. This allows for predicting the location of a potential source of explosives in the next time step. Applying this approach might therefore reduce error rates and respectively wrong alarms (false positives) and help security staff to focus and smaller areas in a given space.

For a first application of our work we chose Munich central station, which processes about 450'000 travelers per day, not including people that use the station as a shortcut between two main streets or enter for shopping. Thus, we are operating in a highly heterogeneous environment with a high passenger density. We consider a special degree of uncertainty in our sensor network in order to keep the analysis realistic and make the approach applicable for the provider. The main aim of the approach is, as mentioned before, not only the detection of the presence of an attacker at a given time within the station, but the position in the grid at a given time and in the next time step.

In order to tackle this complex problem, we chose to capitalize on the recent advances in image classification with convolutional neural networks in the following way. We interpret the measurements of each sensor at their location as gray-scale pixel in an image. To this end we arrange the detectors in a square grid and interpret the probability of detection as gray-scale pixel. Assuming a total number of 100 sensors leads to a $10 \times 10$ input picture. This is based on the fairly realistic assumption, that the conviction behind a possible detection is related to the distance, making nearby explosives easier to detect. One important question that needs to be solved is how to distribute those sensors on the physical train station. Whichever scheme is being used, should maintain the distances between sensors as much as possible. One possibility would be placing a matching grid over the station layout and making sure one sensor is placed in each grid cell, which was also applied in the current state of the project. This might not always lead to sensible or practical results, which therefore needs to be studied further.

We furthermore decided to not just simply take a single image, but instead consider time series of these, to introduce a sense of movement of the attacker. This means that our system can not only draw inference on the readings of the sensors, but also of the pattern changes over time. This requires to couple the convolutional layers with recurrent layers to capture correlation between various time steps.

Since we want to study the problem in the regime of supervised learning, we decided to label each time series with the exact position of the attacker at the time step after the last one in the series or with an additional label interpreted as false alarm. Following the classical approach for classification with
neural networks, this implies having a softmax layer as a final layer. Even just assuming a 10 × 10 grid for the train station leads to 101 labels. In order to improve interpretability of the results, we order the first 100 labels in grid form and convert the probabilities to grayscale pixels, thereby creating a heat map for the location of the attacker at the next time step.

The proposed approach was already implemented and is currently in a test-phase. Therefore, we present it in a working paper for this conference and do not present results in the article.

II. LITERATURE REVIEW

Combining the input of several sensors to detect a signal is a well studied problem. Traditionally they are being approached by techniques based on bayesian formulations [9] and the Neyman-Pearson criterion [21]. It can be placed in the general setting of data fusion and has been studied in a variety of settings [26], [22], [13], [11], [6]. One downside to these approaches is that they require detailed knowledge of all involved probabilities. Even though there are a few papers dealing with this problem [1], [4], [19], they are by far outnumbered. A few papers even apply neural network to the data fusion problem [25], [24], [5], [2]. We intend to capitalize on the recent improvements in neural network to circumvent the problem completely by using the image classification capabilities of convolutional neural networks. Recently there have been several articles suggesting structures similar to ours for video classification [7], [23], [20], [14], [10], [15], [17], [16], [18].

III. MODEL STRUCTURE AND TRAINING ALGORITHM

In order to do the classification from a time-series of pictures we follow a similar structure to [7]. We start with convolutional layers, followed by Gated Recurrent Units and finish with a softmax layer, as can be seen in Fig. 1. Each component will be described more precisely in this chapter.

A. Convolutional and Pooling Layers

The first set of layers is an alternating sequence of convolution and pooling layers, first suggested in [8]. The idea of convolutional layers is to calculate convolutions of a certain number of filters with the input image. Using the same filter for the entire image implies a sense of non-locality, which is then further increased by the accompanying pooling layer. In our case, after the linear combination of the filter with an equally sized part of the image is calculated, it is passed through the non-linearity, in our case a rectified linear unit. This produces one pixel in the new image, whose neighbours are then calculated by shifting the filter across the input image, keeping it still overlapping with the previous position. Once this is done, the new image is passed into a max-pooling layer, in which a square of neighbouring points is replaced by a single pixel, which contains the maximum intensity of the group. This process creates a set of new, smaller images - one for each filter, which are fed into the next layer.

B. Gated Recurrent Units

The Gated Recurrent Units (GRUs) [3] have recently started to replace Long-Short-Term-memory units (LSTM), as they achieve comparable performance with less parameters. A representation can be found in Fig 2. The main component is the convex combination of the previous hidden node at time \( t - 1 \) and a temporary internal representation of the input at time \( t \), that also contains some information of the previous time step. The amount of previous hidden information that enters into the internal representation is learned by the system and depends on the new input as well as the old hidden variable. The same is true for the weights in the convex combination. These parameters essentially determine how much of the past information is remembered. The formulas for this can be seen in Figure 3.

\[
\begin{align*}
  r_t &= \sigma(x_t W_{xr} + h_{t-1} W_{hr} + br) \\
  z_t &= \sigma(x_t W_{xz} + h_{t-1} W_{hz} + bz) \\
  \hat{h}_t &= g(x_t W_{xh} + (r_t \odot h_{t-1}) W_{h} h + b_h) \\
  h_t &= (1 - z_t) \odot h_{t-1} + z_t \odot \hat{h}_t \\
  \sigma(x) &= \frac{1}{1 + e^{-x}}
\end{align*}
\]

Fig. 2. The basic structure of a Gated Recurrent Unit. \( r \) and \( z \) capture how much of \( h \) and \( \hat{h} \) are being used.

Fig. 3. The formulas behind the GRU. \( \odot \) represents elementwise multiplication.
The formulas behind the Softmax Layer. $z_j$ is the inflow to the $j$th unit, $y_j$ its outflow.

C. Softmax Layer

The final layer then needs to map the obtained features to a probability distribution over the labels. To this end, we follow the traditional approach of including a softmax layer. In this layer, each obtained value is passed through an exponential and is then renormalized, so they all sum to one. This also implies that the value of each node depends on all the others. The formulas can be found in fig 4. Typically, this distribution is then used to sample a label from. In our case, it is converted to a heat map for the attacker position, as it is of more value.

D. Adam

To avoid some of the typical problems in selecting the hyperparameters to ensure good training behaviour, we chose to replace plain back-propagation by ADAM [12], which keeps a running estimate of first and second moments. These are then used to adjust the updates. This procedure is reported to work very well in practice and is particularly suited for problems with a large number of parameters and samples, as can be expected here. The details can be found in algorithm 1.

Algorithm 1 ADAM

Require: $\alpha$: Stepsize
Require: $\beta_1, \beta_2 \in [0,1)$: Exponential decay rates for the moment estimates
Require: $f(\theta)$: Stochastic objective function with parameters $\theta$
Require: $\theta_0$: Initial parameter vector
$m_0 \leftarrow 0$
$v_0 \leftarrow 0$
t $\leftarrow 0$

while $\theta_t$ not converged do

t $\leftarrow t + 1$
$g_t \leftarrow \nabla f_t(\theta_{t-1})$
$m_t \leftarrow \beta_1 \cdot m_{t-1} + (1 - \beta_1) \cdot g_t$
$v_t \leftarrow \beta_2 \cdot v_{t-1} + (1 - \beta_2) \cdot g_t^2$
$\hat{m}_t \leftarrow m_t / (1 - \beta_1^t)$
$\hat{v}_t \leftarrow v_t / (1 - \beta_2^t)$
$\theta_t \leftarrow \theta_{t-1} - \alpha \cdot \hat{m}_t / (\sqrt{\hat{v}_t} + \epsilon)$
end while

return $\theta_t$

IV. CONCLUSION

In the ongoing research project presented at hand we developed a learning environment to use sensor data for predictive analysis of attacker movement. In order to tackle this complex problem, we capitalize on the recent advances in image classification with convolutional neural networks. Here, the analysis focuses on time series in order to consider movement within the grid. Since we want to study the problem in the regime of supervised learning, we decided to label each time series with the exact position of the attacker. In the given time of paper submission we were not able to analyze and present first results. However, the approach is a promising first step in the solution of regulatory shortcomings in current technologies for terror prevention. Given that we have precise probabilities for the simulation environment, that produced the first samples of test data, we will be able to match our approach against the bayesian formulation based traditional approaches. We hope to come close to those results, even though information is not used in our approach. Obtaining a sufficient amount of data for this parameter-rich model is not expected to be a problem, since even for a limited number of ultimate goals and entrances there are many different paths the attacker could choose. By also considering subpaths for the 5 frame window, and by varying obstacles the possible number of samples can be scaled to appropriate quantities. Since we based the approach on a real environment, this allows, especially in further steps of the project, for implementing real circumstances and measures for a more realistic view on conditions and results. This will also give the opportunity to experiment further with the sensor placement on the real map and to see how this impacts on the precision.

ACKNOWLEDGMENT

The authors would like to greatfully acknowledge the support by the BMBF Germany, Project RE(H)STRAIN.

REFERENCES