# Deep Learning based MURA Defect Detection

Ramya Bagavath Singh [1], Gaurav Kumar [1,*], Gaurav Sultania [1,*], Shashank Shrikant Agashe [1,*], Priya Ranjan Sinha [1], Chulmoo Kang [2]

[1] Samsung Semiconductor India R&D, Samsung Electronics, Bengaluru, India
[2] Mechatronic R&D Center, Samsung Electronics, Hwaseong, South Korea

## Abstract

**INTRODUCTION**: MURA defects in LED/LCD panels are one of the most challenging defects for Automatic Defect Classification and Localization (ADC) due to their extremely low contrast when compared with the background. Manual detection is subjective, error prone, very tedious and time consuming. Even when the type of MURA defects can be ascertained manually, the exact bounding box for defect is hard to determine. Various heuristic based image processing techniques have been applied giving sub-optimal accuracy over generic datasets.

**OBJECTIVES**: The primary objective of this paper is to check whether the state of the art DL (Deep Learning) network for general object classification and localization (MSCOCO PASCAL VOC etc.) can be applied successfully for MURA Defect Classification and Localization.

**METHODS**: In this paper we present a single DL pipeline for classification and localization which for the first time is applied for MURA defects. Naive DL network - Single Shot multi-box Detector (SSD, pre-trained on ImageNet) was not sufficient to give a good F1 score because of the nature of the defect. Accuracy improved a little after applying various DL specific optimization methods such as loss function optimization, network optimization etc. Utilizing the knowledge from MURA domain for data augmentation, like filtering based on image capture wavelength etc. improved the results significantly.

**RESULTS**: Using optimization techniques that are from both DL domain as well as specific to MURA domain, we show improvement in the accuracy of the base DL pipeline from ~30% to ~80%. Minimum heuristics were used to define the pipeline so that it can easily adapt to any new MURA dataset. The paper shows the importance of domain specific pre-processing steps for the designed network in case of MURA defects.

**CONCLUSION**: Using DL, MURA classification and localization had not been tried before. For the first time we demonstrated results for both classification and localization of MURA defects using state-of-the-art DL network with F1~80%. We also conclude that state-of-the-art network for general object detection can be reused with the help of Transfer Learning (TL) concept and fine-tuned with MURA domain specific optimizations mentioned in paper for optimal performances in MURA domain.

---

*Corresponding author Email: gaurav.k4@samsung.com, gaurav.sl@samsung.com, shashank.a@samsung.com

# 1. Introduction

MURA (Japanese origin) also called blemish or stain defect is very common in any panel manufacturing (OLED/LED/LCD etc.). The defect has peculiar characteristic of having extremely low contrast with the background, making it very difficult to see through naked human eye. MURA defects can range from very small (few pixels) to very big (almost covering entire panel). Due to their low visibility, they are sub-classified mainly according to the backend manufacturing process information (correlation to backend manufacturing process) rather than visual information obtained from defect. Hence, defects from different MURA sub-classes can look similar visually. Moreover depending upon the panel manufacturing process, MURA defect classes can differ and no standard MURA defect classes exist. This also makes it difficult for any existing generic classification and localization pipeline to perform robustly for diverse datasets and hence domain/dataset specific optimizations have to be applied. Nevertheless, the correct classification and localization of MURA defect classes have significant monetary impact for panel manufacturing process by means of reducing root cause analysis time and increasing overall yield due to the correlation to manufacturing mentioned above.

Typically for detecting MURA defects for OLED/LED/LCD panels, images are captured in different wavelengths of light and then passed on to a learning system for automatic detection and classification. This is done to make the defects more visible in some cases; but it's still very difficult to differentiate between the defect and background with naked eye. Illustrative examples of a few types of MURA defects are given in [1]. As can be seen therein, the defects appear as low-contrast, non-uniform brightness regions and they are typically larger than a single LCD pixel. Additionally, multiple defects of different types can occur simultaneously as well as multiple times in a single image. So correct classification and localization of every defect in the panel is very important to know all the root causes responsible for defects. In our dataset there are 4 types of MURA defect (Note due to confidentiality we cannot show the real defect images here but described defect by text below). Type 0 and 1 are similar looking defect known as patch defect usually small in size and differ only by the location where they occur in panel. Type 1 defect only occurs at edges, whereas Type 0 can occur anywhere in panel. Type 2 defect is big defect also referred to as skin peel or skin rash defect due to weak visual similarity to skin peel or skin rash. Type 3 is very small defect known as spot or point defect and they occur with very high probability density.

Recently Deep Learning (DL) based pipelines have become state of the art for various recognition and object detection tasks. Naturally DL can be thought as a promising approach for MURA defect classification and localization as well. State of the art DL classification and localization pipelines however are trained and tested over normal objects of day to day occurrence which have some visual constraint over shape and size like bus, train, people etc. As discussed above, it is not true for MURA defects; their classification is more correlated with the manufacturing process compared to their appearance. So directly applying standard DL methods tor MURA is unlikely to give good results.

Table 1. Custom F1 Score

| Actual Class | Predicted Class | | | |
| --- | --- | --- | --- | --- |
| | Type A | | Other | |
| IOU | >0.5 | <0.5 | >0.5 | <0.5 |
| Type A | $TP^A$ | $FP^A$ , $FN^A$ | $FP^{other}$ , $FN^A$ | $FN^A$ , $FP^{other}$ |
| Other | $FP^A$ , $FN^{other}$ | $FP^A$ , $FN^{other}$ | $TP^{other}$ | $FN^{other}$ , $FP^{other}$ |

In this paper we present a DL pipeline inherited from state-of-the-art DL pipeline for classification and localization of normal objects (from MS COCO, PASCAl VOC 2012, 2007 etc.) and optimized with domain specific knowledge to get high accuracy for MURA defects. For final metric we define custom F1 score and report it per defect as well as for overall test set. Table 1. show the custom F1 score metric used in this paper, where Intersection of Union (IOU), True Positive (TP), False Positive (FP), True Negative (TN) and False Negative (FN) are shown in abbreviated form. Type A is any MURA defect class for which F1 score is calculated. Overall F1 score can then be simply calculated by summing TP, TN, FP and FN over all MURA defect classes. Our dataset consists of 4 MURA defect classes. We have deviated from the standard mean average precision (MAP) reported in related literature for quantifying the quality of DL pipeline because the manufacturing yield is also dependent on recall; bad recall can lead to un-necessary tuning of process step parameters thus affecting the manufacturing yield.

Lastly, the contribution are twofold:-

- We present the first application of DL for MURA defect classification and localization. We also define minimum heuristics (no tunable thresholds) pipeline which is much easier to adapt for newer MURA dataset (OLED/LED/LCD). Whatever minimum heuristic is defined comes from inspection process setup (specifically the different wavelengths used for imaging) which can be common across different MURA dataset.
- We present technique for re-using state-of-the-art DL pipeline for classification and localization trained on normal objects. We rely heavily on Transfer Learning (TL) concept and process by fine-tuning the inherited DL pipelines for our dataset. Some part of domain specific optimization techniques described here can be applied to any DL pipeline. So as state-

of-the-art DL pipeline improves for normal objects, they can be directly plugged in for improved accuracy for MURA.

Throughout this paper, although we disclose results for real MURA defects, we will use publicly available images for illustration purposes. This is done in order to preserve the confidentiality of sensitive data. The rest of the paper is organized as follows. Section 2 gives overview of the related work in the field. Section 3 contains detailed explanation of proposed pipeline while the results are presented in section 4.

## 2. Related work

### 2.1. Literature survey for MURA defect inspection

There are several electrical and vision based inspection techniques available for MURA defect inspection [2]–[8].

In TFT-LCD the voltage-imaging technique measures the characteristics of a LCD array by directly measuring the actual voltage distribution on the TFT pixels. However, probes used for voltage measurement must be separately designed for each panel configuration. In vision-based techniques, Song et al. [2] developed a wavelet based method to detect the MURA defects in low-resolution LCD images that involve non-textured surfaces. Lu et al. [3] applied the Independent Component Analysis (ICA) to detect defects in patterned LCDs. These approaches define hand-crafted heuristics and thresholds which had to be separately designed for different MURA defects. To overcome this limitation traditional machine learning (ML) approaches have also been applied. Liu et al. [4] used the Locally Linear Embedding (LLE) to extract image features and then applied Support Vector Machine (SVM) for classification without localization. To perform localization Kim et al. [5] used adaptive multi-level defect detection and probability density estimation for TFT-LCD inspection. Lin et al. [6] presented an image processing method for defect detection in TFT-LCD images and used genetic algorithm (GA) for adjusting heuristics automatically. Ngo et al. [7] also presented an automatic detection method for MURA by accurate reconstruction of the background by training separately on the background but using test set images of MURA. In non-ML based method, Du-Ming Tsai et al. [8] used Fourier transform based technique to remove the repeated patterns in background and then used adaptive threshold to perform defect segmentation. These traditional ML and non-ML methods, though successful in some cases, fail to adapt successfully to more generic datasets.

DL techniques have also been applied for MURA defect classification. Hua yang et al. [1] applied TL and deployed an Extreme Learning Machine (ELM) for online MURA defect classification with impressive results. DL methods for both classification and localization of defects have been applied for defects other than MURA. Liu Ri-Xian et al. [9] applied Deep Belief Network (DBN) as goodness of fit for defect identification in capsule and solar cells. Adaptation of DL in MURA domain is still limited mainly due to scarcity of public datasets. Even when in-house dataset is available, the number of training images are usually less as individual images have to be manually labelled. Additionally, due to low contrast of defects, manual labels are also not very reliable. Our work overcomes these limitations by utilizing state-of-the-art DL pipelines trained for normal objects and modifying them appropriately for MURA datasets containing small number of images. The authors are of the opinion that this will facilitate widespread usage of DL in MURA domain.
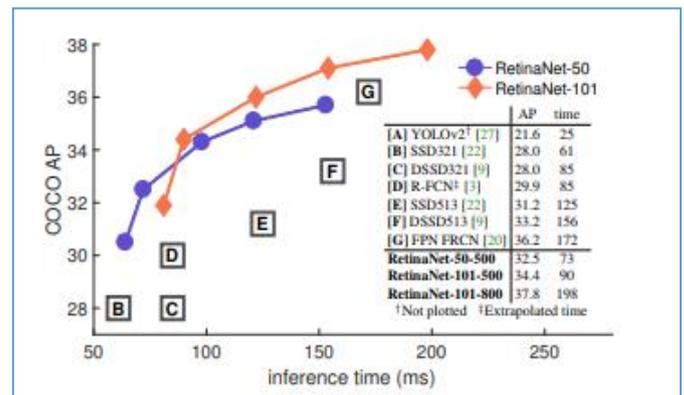


**Figure 1.** Literature Survey for state of art DL for normal object detection [12]. (2017. Recent new state of art DL networks has come but they can fit directly here in this paper)

### 2.2. Literature Survey of state of art DL pipeline for normal object detection

For normal object classification and localization, DL techniques have outshined other techniques in terms of performance. MS COCO, Pascal VOC 2012, 2007 are some of the challenges in this area where DL techniques have consistently occupied top position in recent years. Literature survey reveals two major variants of DL pipelines being used in this area: single stage and two stage. In single stage classification and localization happens in a single pipeline whereas in two stage variant two separate pipelines are used. First pipeline gives object proposals and second pipeline performs classification and localization over the predicted proposals. The pipelines can be trained simultaneously or separately. Fig. 1 shows a consolidated view of speed vs. accuracy for different pipelines. An apparent trade-off is visible for the two categories of performance evaluation. In next section we will present the base DL network pipeline from which our final DL pipeline (working on MURA defect dataset) is inherited.

# 3. Proposed DL pipeline for MURA

## 3.1. Choosing Base DL Pipeline
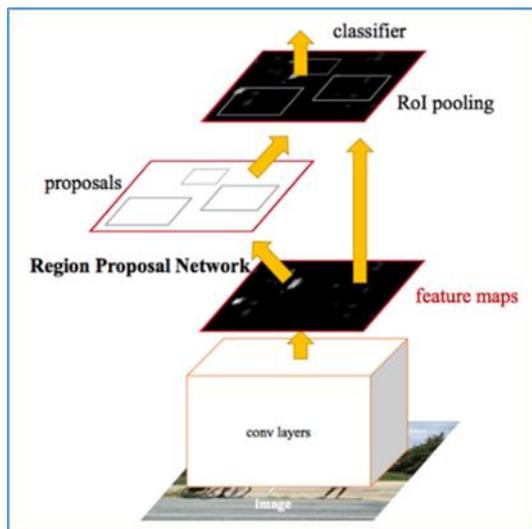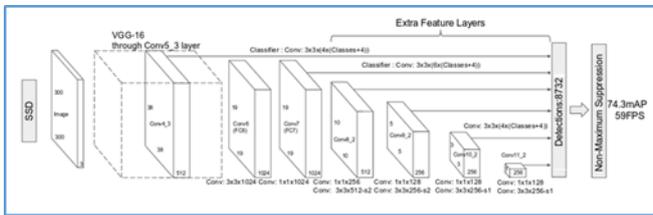




**Figure 2.** Single stage and Two stage DL networks

We tried both single stage and two stage DL based classification and localization pipelines to choose the base network. In single stage we tried Single Shot Detector (SSD) which also encompasses YOLO [10] and for two stage we tried variants of Faster Region Convolutional Neural Network (Faster RCNN). Fig. 2 illustrates the two said pipelines. Faster RCNN performed poorly than SSD in default training configuration particularly because of the low performance of Region Proposal Network (RPN). Overall F1 for Faster RCNN was ~10% while that for SSD was ~30%. As MURA defects don't have well-defined boundary that segregates them from the background, RPN fails to learn the foreground object pattern and gives low quality proposals which affect the overall accuracy of the Faster RCNN pipeline. Comparatively, the dense proposal matching in case of SSD performs better. Training with only MURA dataset of 344 images was insufficient especially given the depth of the pipelines. Typically, dataset of the size of ImageNet is required for training such deep networks. So we utilized the TL concept: for feature generation in both the pipelines we used the pre-trained weights from detection pipeline trained on ImageNet. Only the last block of feature network was fine-tuned over our dataset. We tried with pre-trained weights of VGG16 [13] and RESNET 51 [14], both giving almost equal score with RESNET performing marginally better (~1%) but at the cost of

increased training time. So for our base model accuracy we selected SSD with pre-trained VGG 16 [14] which give combined F1 score of ~30%.

## 3.2. Modified DL Pipeline for MURA inspection

To increase accuracy of base network we applied many optimization strategies which can be broadly divided into network specific and domain specific. In next paragraph we discuss network specific optimizations followed by domain specific optimizations in subsequent paragraph. Note that network specific optimization is specific to state-of-the-art network chosen as base (SSD for this paper). Domain specific optimizations are independent and can be applied to any new DL pipeline.

For network specific optimization we modified two things:-

- Loss Function Optimization:-

  The SSD pipeline uses multi-box loss for training. In multi-box loss the loss gradients are applied only to the overlapping boxes (the proposal box having greater than 0.5 overlap with GT) and equal number of non-overlapping boxes randomly chosen from all the proposals. This accounts for less than 1% of boxes being trained per batch. This causes training to be slow and also many of the boxes remain un-trained even after the training process (mainly due to low training set size of about 344 images compared to millions of images in ImageNet). We changed the multi-box loss with weighted loss where all the proposal are simultaneously trained with loss gradient, which get proportionately divided between overlapping and non-overlapping boxes as per the ratio of their count. This results in improvement of F1 score on test set by ~10%.

- Generic network optimization:-
  Further improvement of the base network was performed by employing following techniques: regularization using as dropout, adding batch normalization to control the variation between layers and augmenting training dataset to 4x by using generic modifications such as image flip. These actions resulted in improvement of test F1 score by additional ~10%.

For domain specific optimization we perform following:-

- Image pre-processing:-

```
def standardize_adv(im, gt=None, newshape=None):
    avg = np.mean(im)
    std_dev = np.std(im)
    im[:,:,0] = (im[:,:,0] - avg)/ (0.5*std_dev)
    im[:,:,1] = (im[:,:,1] - avg)/ 1*std_dev
    im[:,:,2] = (im[:,:,2] - avg)/ 2*std_dev
    return im, gt
```

**Figure 3.** Pseudocode for modified standardization used as pre-processing

We tried pre-processing steps to specifically increase contrast between defect and background. Fig. 3 shows the pseudocode for modified standardization which was found to increase the contrast between the defects and background the most. This preprocessing method increased the F1 score by additional ~10%.

- Domain specific data augmentation:-

In our dataset defects of type 0, 1 and 3 (especially type 3) are small defects compared to panel image size whereas type 2 is much bigger defect. So we define a pre-processing step exclusively for small defects which we call as crop and combine as illustrated in Fig. 4. We perform ordered crop and during training we supply only those crops which contain defect. However during testing we supply all the crops in ordered fashion and perform concatenation of the result. This technique increases the F1 score of smaller defects (especially for Type 3 which was more than ~10% increase) but decreases score for the bigger defects. Overall Score increased marginally by ~5% as our dataset contained more number of smaller defects.
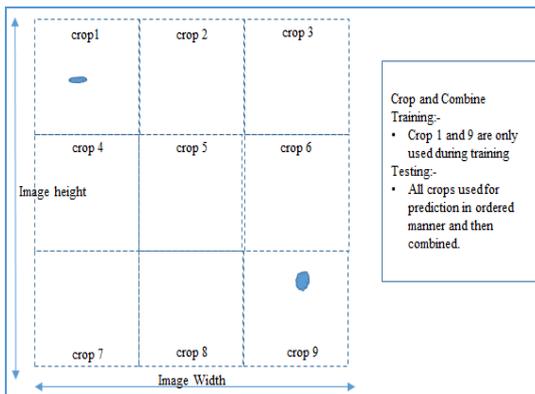


**Figure 4.** Crop and Combine data augmentation technique

- Ensemble Network: -

Due to large correlation of MURA defect classes with background manufacturing steps, defects from different classes may appear similar. We observed this empirically as well; the trained DL network gets confused with similar looking defects. So we divided the network into separate networks; each network detecting different class of similar looking defect. Also as crop and combine technique (figure 4 above) can only be applied to small defects, we trained separate networks for bigger and smaller defects as well. We trained an ensemble of three networks as shown in Fig. 5: first network trained for type 0 and 1, second network trained for type 2 and third network trained for type 3. Note that even though type 0 and type 1 look similar, network didn't confuse between them because type 1 only appeared at edges of the panel image. During testing we passed test images to all the networks and consolidated the output. Each network performed better individually (~5% increase) on their specific test sets containing only the defects for which they were trained; however the overall score decreases by ~5% due to increase in false positive cases (especially for the case when test image of defect was supplied to network which was trained for other defect).

- Wavelength based filtering before final prediction (Information specific to Inspection Setup instrument):-

Table 2. Wavelength Filtering

| MURA Defect Class | Wavelength Rule |
|---|---|
| Type 0 | All wavelength except wavelength index 0 |
| Type 1 | Only wavelength index 1 |
| Type 2 | Only wavelength index 2 |
| Type 3 | Only wavelength index 0 |

In our dataset all defects (type 0, 1, 2 and 3) are provided in different wavelength. We empirically learnt from our training history (past train F1 Scores) that for each defect class, input images corresponding to certain wavelengths result in improved performance of the DL network. With this information we created a rule based filter (shown in table 2) using the wavelength of input image and added it just before calculating final metric in our ensemble network setup. Thus, as per the rule in table 2, DL network prediction of type 0 on any input wavelength image other than 0 would be trusted. The resulting ensemble network with this filtering technique increased overall F1 score by about ~30% giving the overall final score of ~80%. This huge increase in score depicts the importance of domain specific knowledge especially in case of MURA defects.
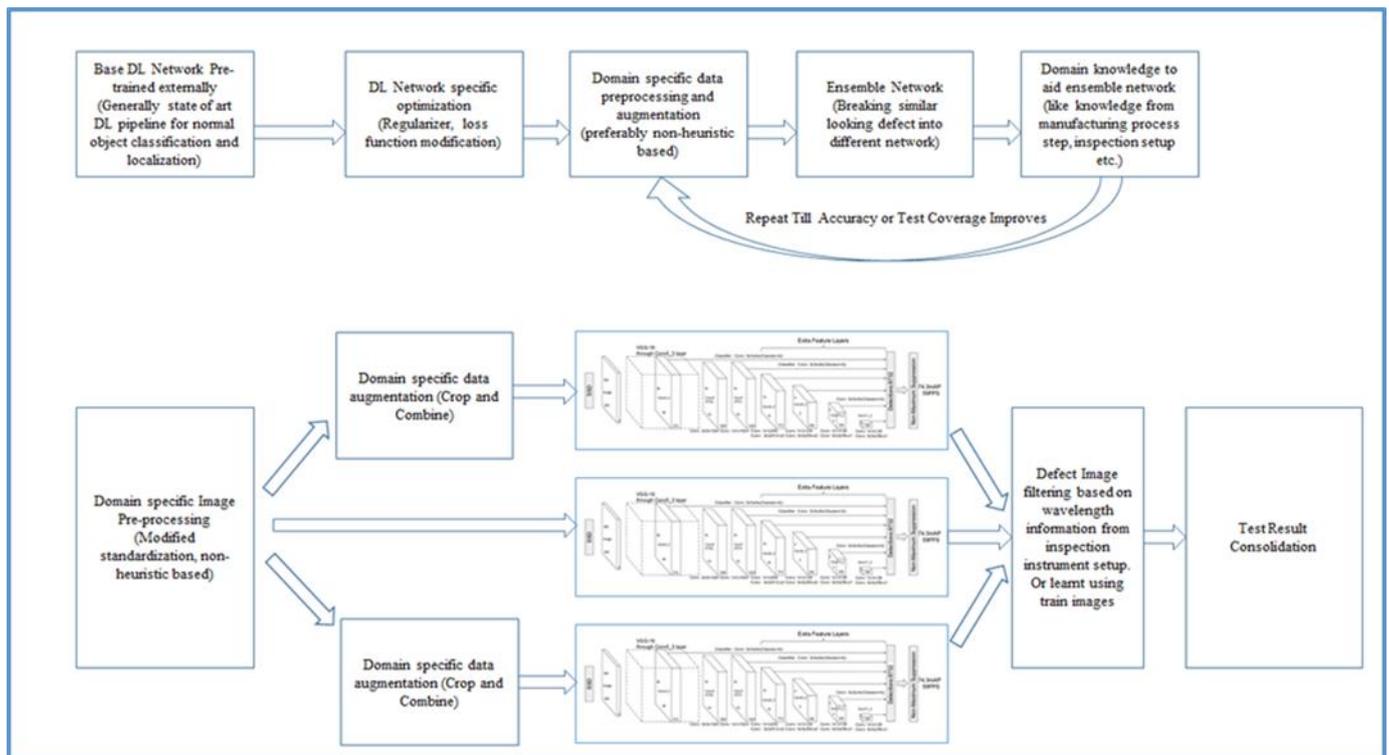
**Figure 5.** Depicting the general optmization strategy followed in this paper (top). Final DL Network used for MURA dataset (bottom).

# 4. Results and conclusion

Our in-house MURA dataset consisted of 344 images as already mentioned. We performed 5 fold cross-validation on our dataset as well as for each fold we ran train and test evaluation for 5 times. This is to ensure that we average out any effect due to random initialization of parameter. Due to large time in training we did not integrate hyper-parameter tuning to further increase the final F1 score. We fixed the training epoch to 100 and saved model state after each epoch. For reporting metric we take mean of the 25 F1 values as well as the standard deviation of the same. Table 3 shows the result summary. One can also see how the F1 score changes with induction of additional analysis methods.

## Table 3. Result table

| Dataset (MURA (1480X720)(344 images) | F1 Score (Table 1) in percentage on 20% test split. Reported = mean score (standard deviation) | Processing Time in milli-sec (ms) |
|---|---|---|
| Base Network (SSD)<br><br>(1 model to train) | Type 0=11.53 (1.32)<br>Type 1=10.71 (0.64)<br>Type 2=83.86 (1.40)<br>Type 3 = 7.24 (2.20)<br>Overall=30.51 (2.66) | Training:- 30000 ms/epoch<br><br>Test:- 50 ms/image |
| ++[a] generic network optimization<br><br>(1 model to train) | Type 0=36.66 (6.32)<br>Type 1=44.87 (2.09)<br>Type 2=74.11 (1.87)<br>Type 3 = 34.17 (9.79)<br>Overall=42.94 (5.26) | Training:- 30000 ms/epoch<br><br>Test:- 50 ms/image |
| ++[a] Image prep-processing<br>(1 model to train) | Type 0=55.89 (3.30)<br>Type 1=57.08 (3.32)<br>Type 2=98.89 (1.53)<br>Type 3=30.61 (4.27)<br>Overall=57.79 (5.12) | Training:- 30000 ms/epoch<br>Test:- 50 ms/image |
| ++[a] Domain specific data augmentation<br>(1 model to train) | Type 0=59.46 (2.58)<br>Type 1=66.42 (2.80)<br>Type 2=82.54 (2.28)<br>Type 3=40.38 (3.42)<br>Overall=61.17 (3.86) | Training:- 30000 ms/epoch<br>Test:- 50 ms/image |
| ++[a] Ensemble Network<br>(3 models to train) | *Type0=73.56 (7.85)<br>*Type1=70.44 (5.60)<br>*Type2=96.76 (5.26)<br>*Type3=66.44 (6.69)<br>Overall=56.18 (4.86) | Training:- 70000 ms/epoch[b]<br>Test:- 150 ms/image |
| ++[a] Wavelength based filtering before final prediction<br><br>(3 models to train) | Type 0=72.68 (3.81)<br>Type 1=99.32 (2.03)<br>Type 2=98.65 (2.11)<br>Type 3=85.88 (3.94)<br>**Overall=81.98 (1.98)** | Training:- 70000 ms/epoch[b]<br><br>Test:- 150 ms/image |

a. Denote increment to previous model state.

b. With multi GPU this can be performed parallel to reduce time further.

*Defect Specific test-set is used for getting the results

We have thus demonstrated results for classification and localization of MURA defects using state-of-the-art DL network with F1~80%, which is the best result of any that are reported for this purpose. The pre-processing steps and the network design employed can form the basis for future work in this field. Due to sensitive nature of MURA data, in this paper we could not add actual prediction result images of our pipeline over test-dataset.

## References

[1] H. Yang, S. Mei, K. Song, B. Tao and Z. Yin, (2018) "Transfer-Learning-Based Online Mura Defect Classification", IEEE Transactions on Semiconductor Manufacturing, vol. 31, no. 1, pp. 116-123.

[2] Y. –C. Song, D.-H. Choi and K.-H. Park, (2006) "Wavelet-based image enhancement for defect detection in thin film transistor liquid crystal display panel", Japanese Journal of Applied Physics, vol. 45, pp. 5069-5072.

[3] C.-J. Lu and D.-M. Tsai, (2008) "Independent component analysis based defect detection in patterned liquid crystal display surfaces," Image and Vision Computing, vol. 26, pp. 955-970.

[4] Y.-H. Liu, Y.-K. Huang and M.-J. Lee, (2008) "Automatic inline defect detection for a thin film transistor-liquid crystal display array process using locally linear embedding and support vector data description", Measurement Science and Technology, vol. 19, 095501.

[5] S.-Y. Kim, Y.-C. Song, C.-D. Jung and K.-H. Park, (2011) "Effective defect detection in thin film transistor liquid crystal display images using adaptive multi-level defect detection and probability density function", Optical Review, vol. 18, pp. 191-196.

[6] C.-S. Lin, Y.-C. Liao, Y.-Long Lay, K.-C. Lee and M.-S. Yeh, (2008) "High-speed TFT LCD defect-detection system with genetic algorithm," Assembly Automation, vol. 28, pp. 69-76.

[7] C. Ngo, Y.-J. Park, J. Jung, R.-U. Hassan and J. Seok, (2017) "A new algorithm on the automatic TFT - LCD mura defects inspection based on an effective background reconstruction", Journal of Society for Information Display, vol. 25, pp. 737-75.

[8] Du-Ming Tsai, Yan-Hsin Tseng and S. K. Morris Fan, (2018) "Defect Inspection of Liquid-Crystal-Display (LCD) Panels in Repetitive Pattern Images Using 2D Fourier Image Reconstruction", International Conference on Autonomic and Autonomous Systems.

[9] Ri-Xian, L & Ming-Hai, Y & Xian-Bao, (2015) "Defects detection based on deep learning and transfer learning", Metallurgical and Mining Industry, 7, pp. 312-321.

[10] Joseph Redmon, Santosh Divvala, Ross Girshick and Ali Farhadi, (2015) "You Only Look Once: Unified, Real-Time Object Detection", CVPR 2015.

[11] Wei Liu, Dragomir Anguelov, Dumitru Erhan, Christian Szegedy, Scott Reed, Cheng-Yang Fu, Alexander C., (2016) "SSD: Single Shot MultiBox Detector", in ECCV 2016.

[12] Tsung-Yi Lin, Priya Goyal, Ross Girshick, Kaiming He and Piotr Dollár, (2017) "Focal Loss for Dense Object Detection", CVPR 2017.

[13] Karen Simonyan and Andrew Zisserman, (2014) "Very Deep Convolutional Networks for Large-Scale Image Recognition", CVPR 2014.

[14] Kaiming He, Xiangyu Zhang, Shaoqing Ren and Jian Sun, (2015) "Deep Residual Learning for Image Recognition", CVPR 2015.