OBJECT RECOGNITION

WITH LIMITED DATA SETS AND COMPLEX ENVIRONMENTS

ECE 6258 – SEMESTER PROJECT

JORDAN LEIKER

CREATING THE NEXT®

PROBLEM STATEMENT

OBJECT RECOGNITION: CURE-OR

PROBLEM STATEMENT

OBJECT RECOGNITION CHALLENGE WITH CURE-OR DATASET

- Develop a model to classify objects in different settings with ‘limited’ training data
  - 10 categories of object
  - 16 categories of distortion
  - 2D backgrounds for training with each distortion
  - 3D backgrounds for testing with same distortions

Figure 1: Example object with 2D backgrounds (image 1 thru 3) and 3D backgrounds (image 4 and 5)

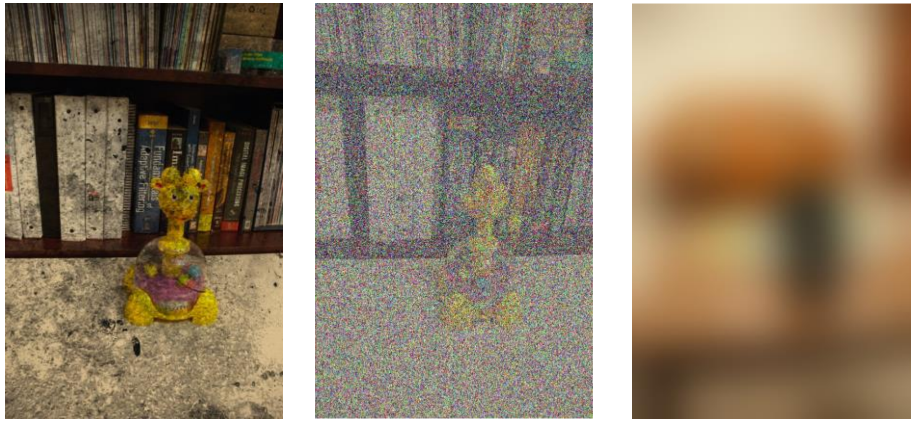


Figure 2: Example distortions

METHOD / ALGORITHM

OBJECT RECOGNITION: CURE-OR

CREATING THE NEXT®

METHOD / ALGORITHM DESCRIPTION

- Machine Learning approach – Keras + TensorFlow backend
  - Keras “ImageDataGenerator” class for image augmentation (artificially increase the size of the training set)
    - Added shear, zooming in/out, horizontal flipping
  - Limited compute resources forced me to use shallow neural nets, smaller layers, small batch sizes, and train on a small number of epochs
- Pre- and post-processing of images critical to performance due to lack of robust CNNs/models – Dippykit + OpenCV
  - Eliminate distortion/noise when possible before predictions
  - Remove as much unnecessary background as possible

ALGORITHM DESCRIPTION

- Two CNN's trained
  - One on category (CNN-A)
  - One on distortion (CNN-B)
- Pre- and Post- processing
  - Image 'normalization'
  - Distortion removal
  - Object location and background smoothing

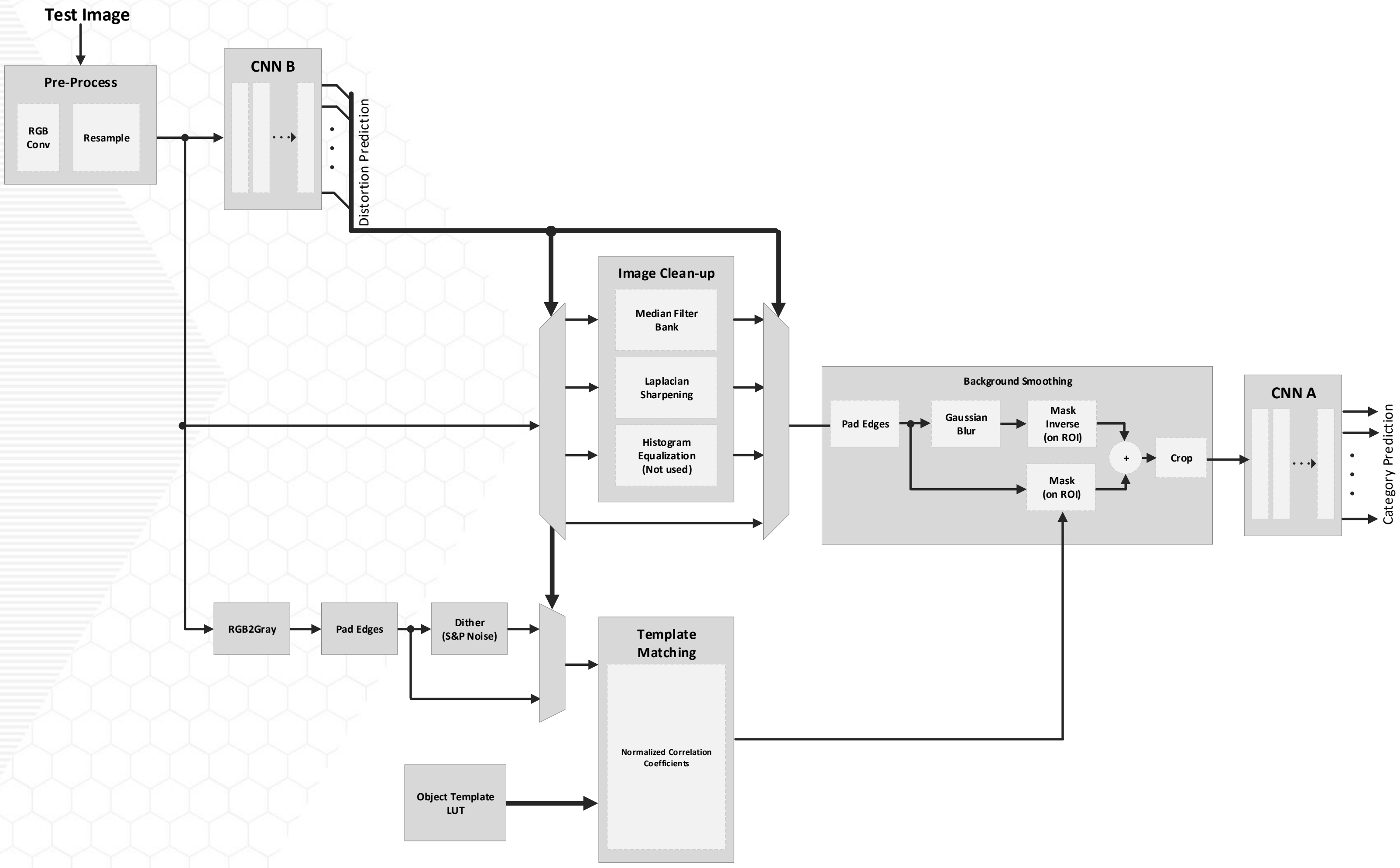
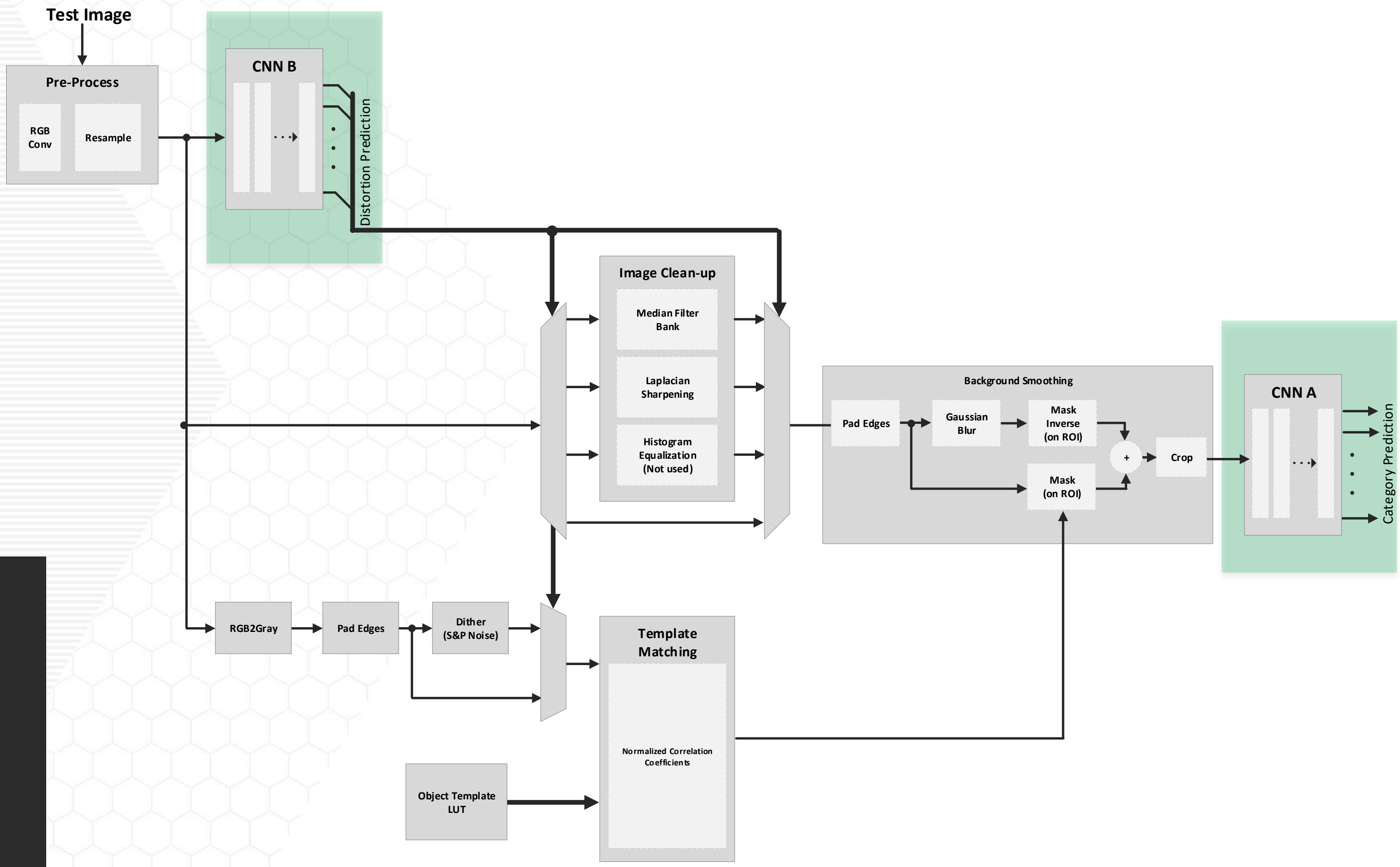


Figure 3: Image Processing Pipeline

ALGORITHM DESCRIPTION – CNN'S

- CNN's are very similar
  - Final dense layer sizes are 10/7 (objects/distort's)
  - Did not train on all data. 'No challenge' and 'resize' distortions left out
- CNNs very shallow, not tuned/trained as much as they should've been
  - Object: ~7 Epochs @ 70% acc, 75% val-acc
  - Distorts: ~5 Epochs @ 98% acc, 99% val-acc

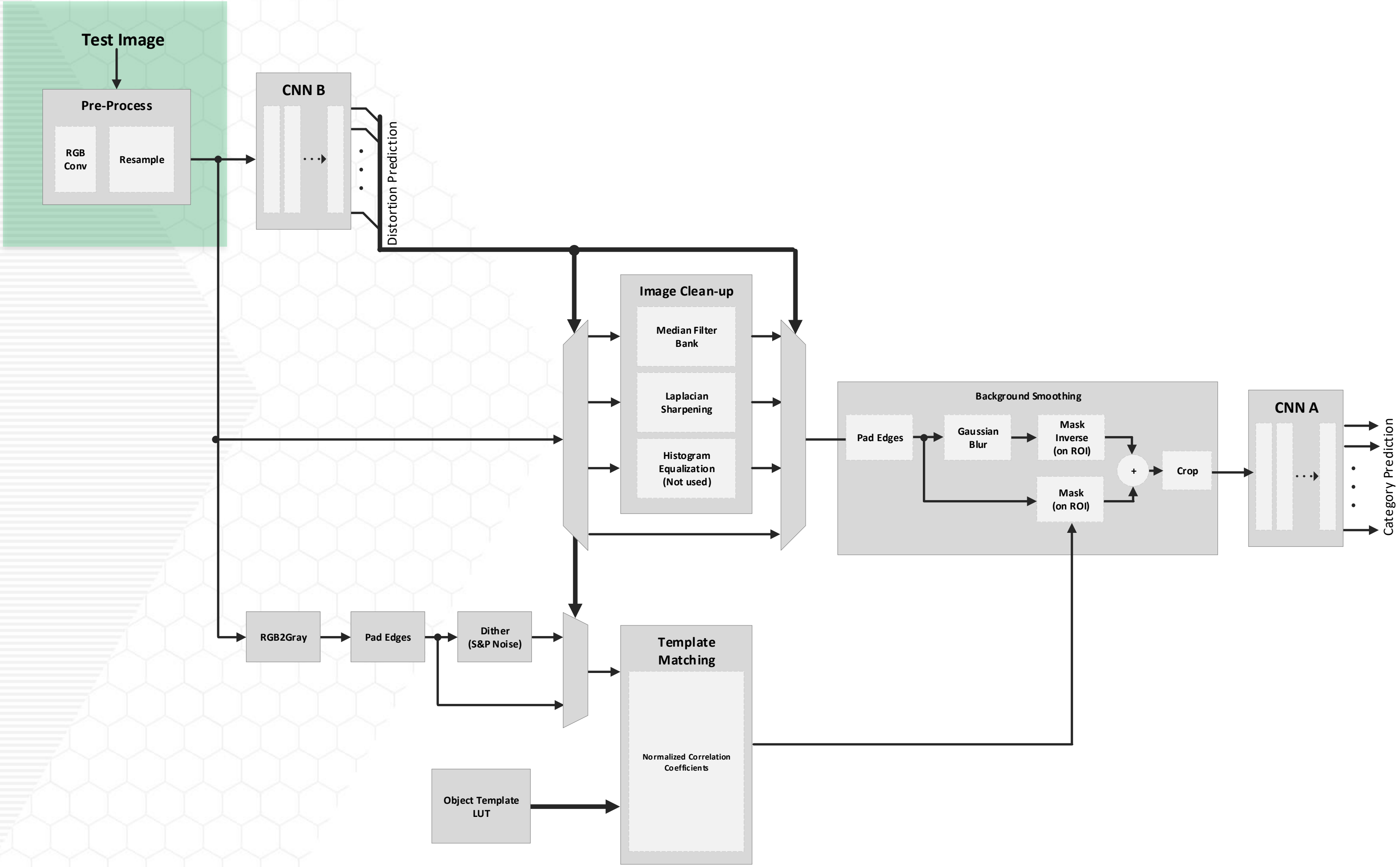
```
model.add(Conv2D(32, (3, 3), padding='same', input_shape=(968, 648, 3)))
model.add(Activation('relu'))
model.add(MaxPooling2D(pool_size=(2, 2)))

model.add(Conv2D(32, (3, 3)))
model.add(Activation('relu'))
model.add(MaxPooling2D(pool_size=(2, 2)))

model.add(Conv2D(64, (3, 3)))
model.add(Activation('relu'))
model.add(MaxPooling2D(pool_size=(2, 2)))

model.add(Flatten())
model.add(Dense(128))
model.add(Activation('relu'))
model.add(Dropout(0.5))
model.add(Dense(10))
model.add(Activation('softmax'))
```



ALGORITHM DESCRIPTION – PRE PROCESS

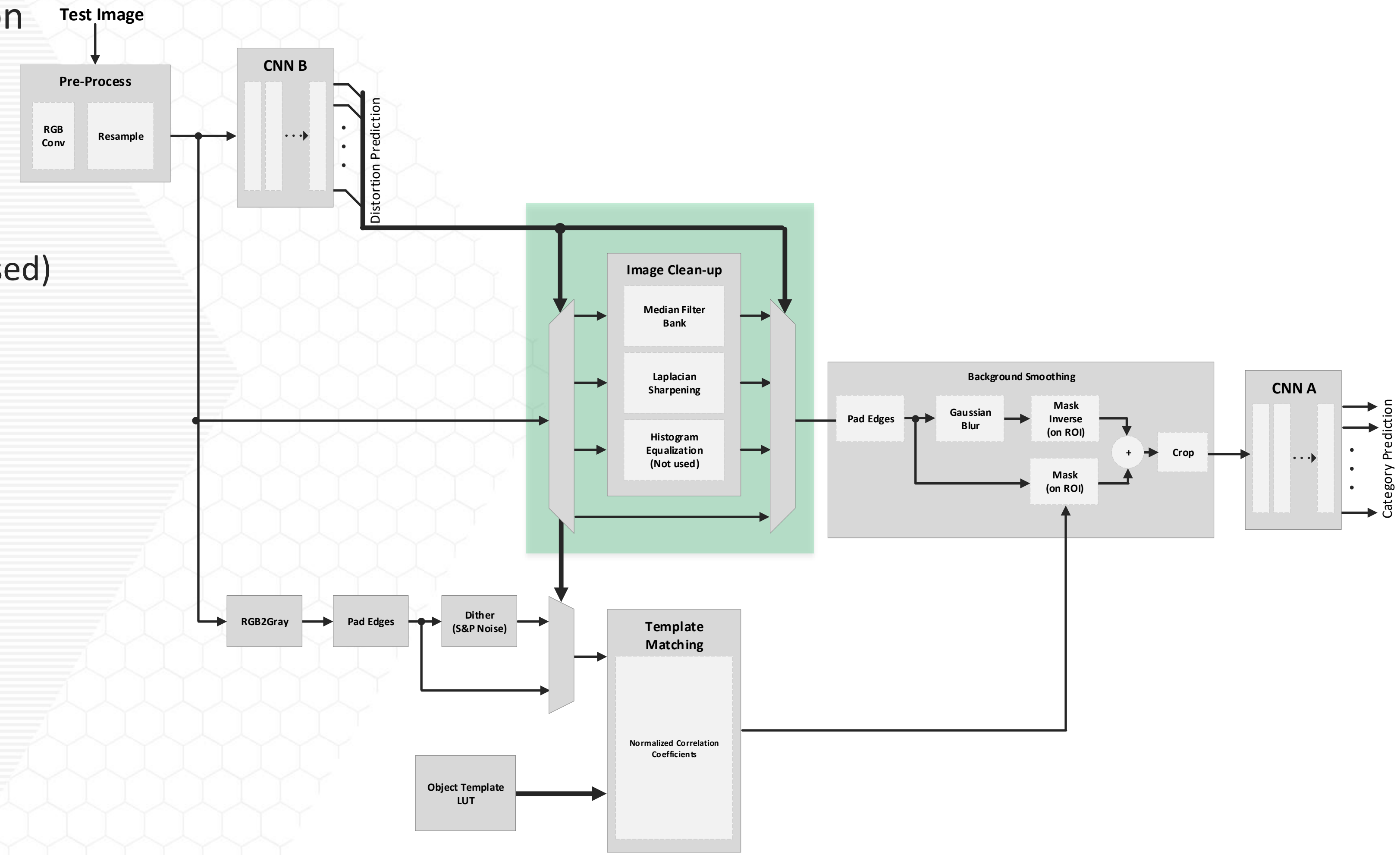
- Pre-process
  - Convert all images to RGB
  - Resample (upsample) images as close to 968x648 as possible
    - Crop images that go slightly over
    - Pad (with reflection) images that go slightly under

- Keras expects standard image
  - Virtually eliminates mis-classification due to 're-size distortion'



ALGORITHM DESCRIPTION – DISTORTION REMOVAL

- Filter / Sharpen based on CNN-B distortion prediction
  - Salt & Pepper -> Median Filter
  - Gaussian Noise -> Laplace Sharpen
  - Saturation/Contrast -> Hist Equal. (not used)
- S&P and Gauss were the only statistically significant mis-identifications



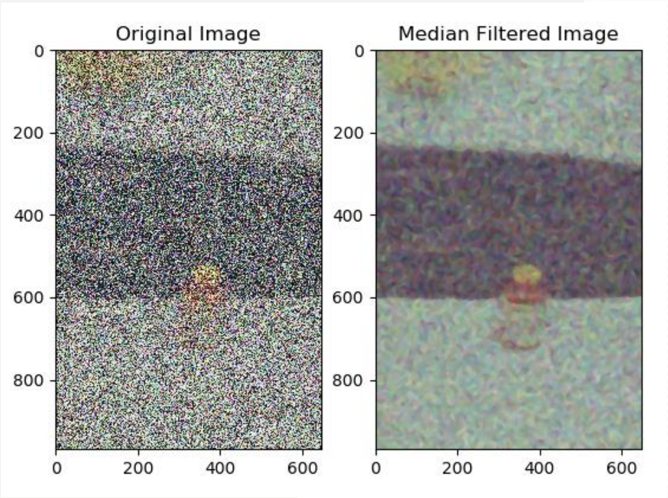
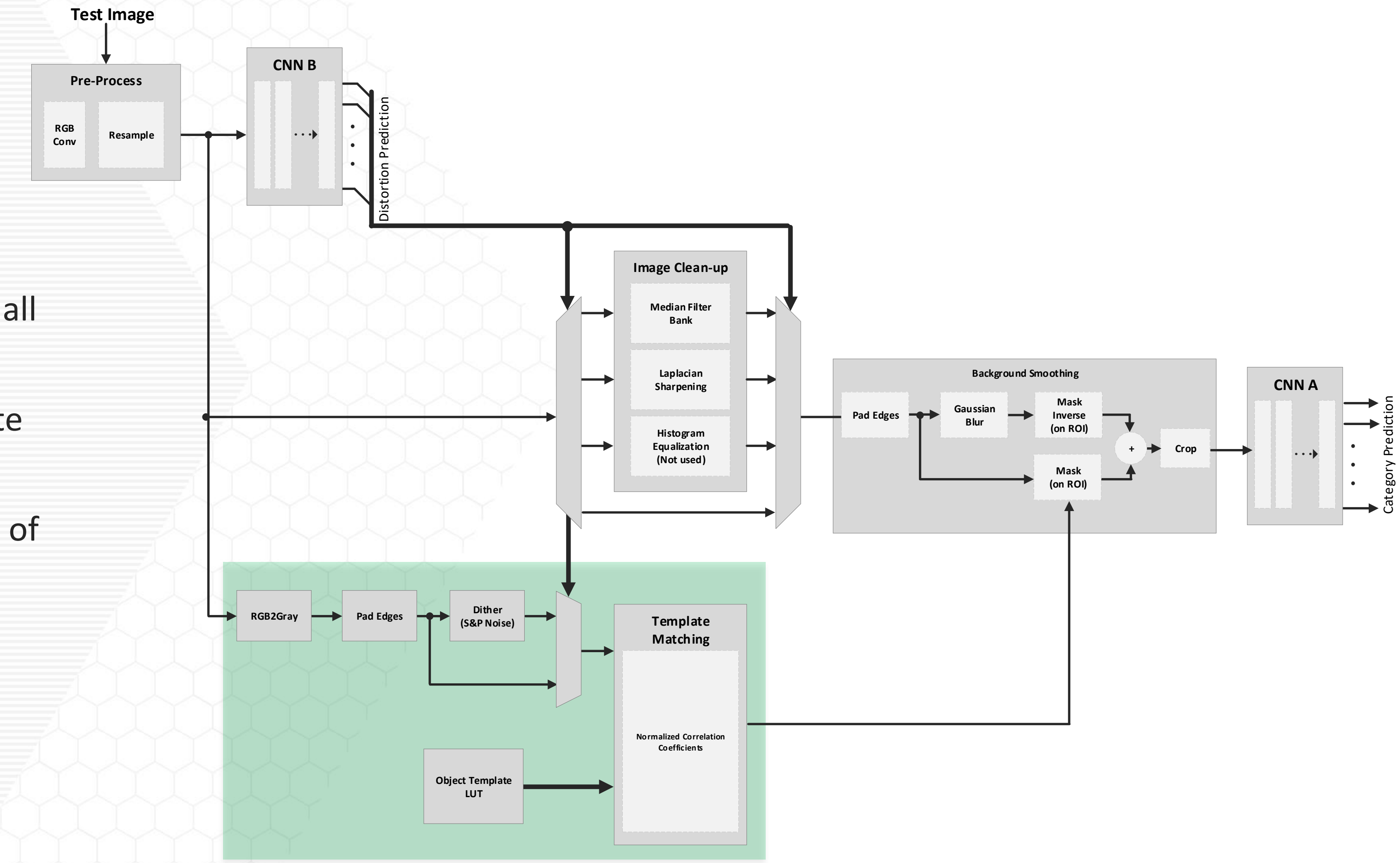


Figure 4: Before and after of Salt & Pepper Cleanup

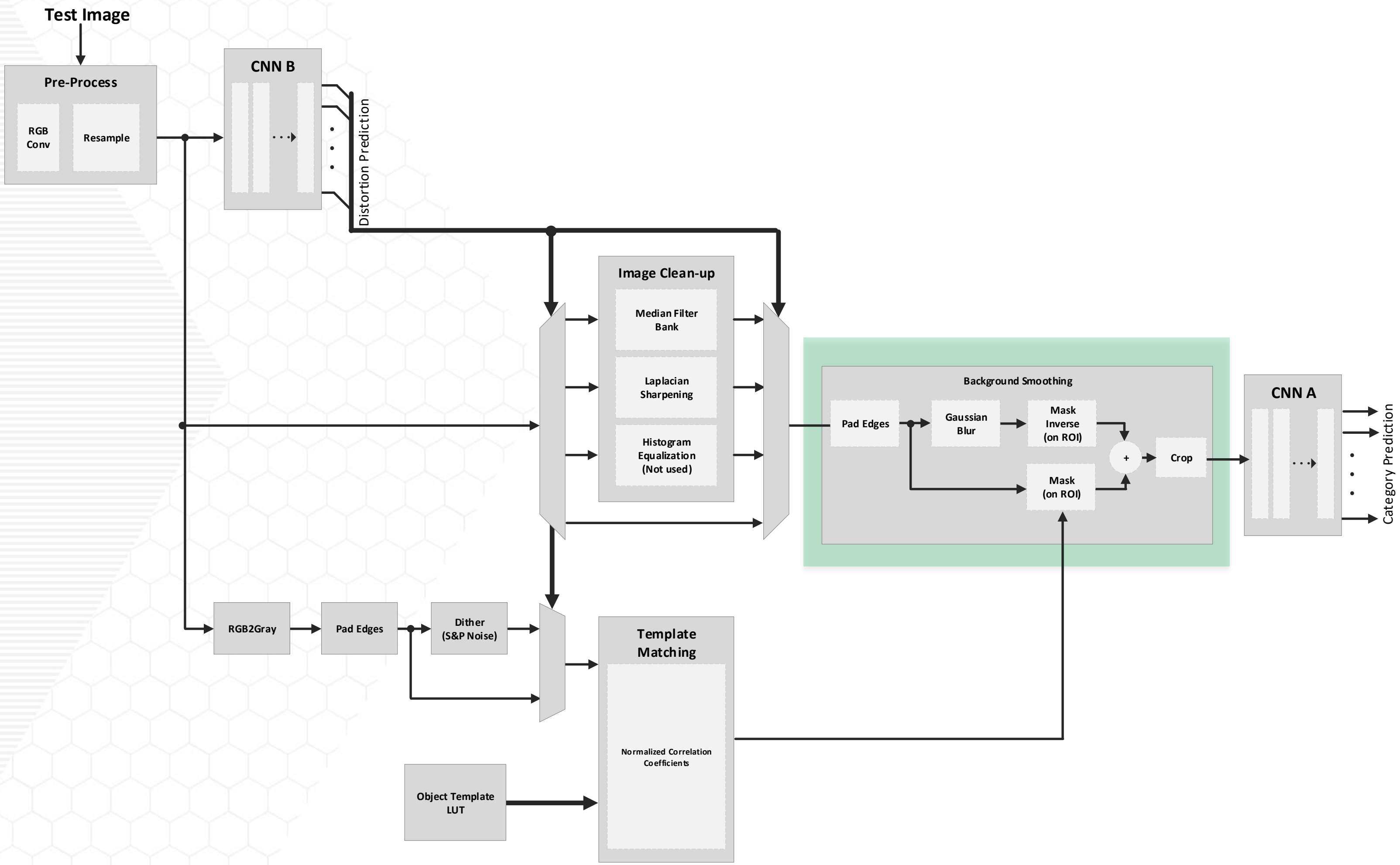
ALGORITHM DESCRIPTION – OBJECT LOCATION

- Object Location and Background smoothing critical for identification over 3D backgrounds
- Template Matching based approach
  - Template was a ‘morph’ of all objects in all orientations
  - Pad images so enlarged ‘morph’ template can slide across full image
  - Dither (add S&P noise) to reduce affects of 3D background during matching



ALGORITHM DESCRIPTION – OBJECT LOCATION CONT.

- Background around object 'smoothed' prior to category prediction
  - Pad the 'cleaned' image so background blur does not effect original image
  - Area around object is masked off (ellipse mask)
  - Remaining background is blurred with Gaussian noise



ALGORITHM DESCRIPTION – OBJECT LOCATION CONT.

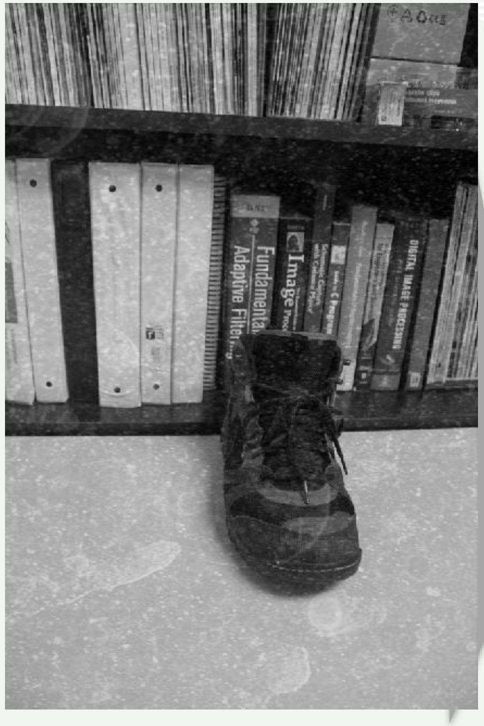


Figure 5: Test image with dirty lens distortion and 3D background

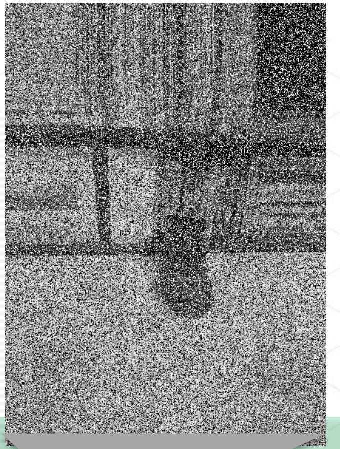
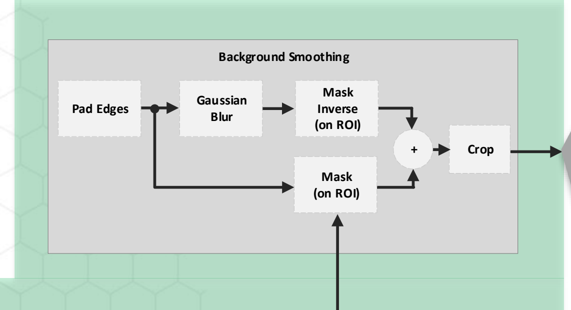
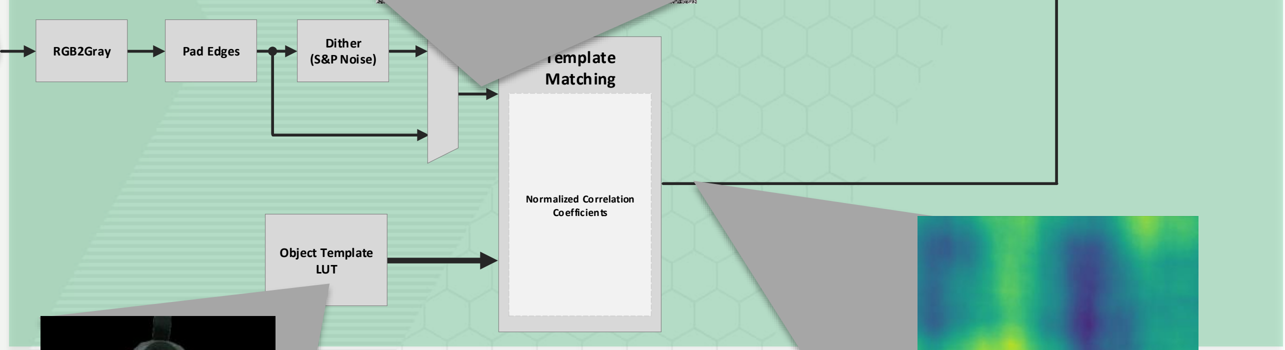


Figure 6: Padded and Dithered test image





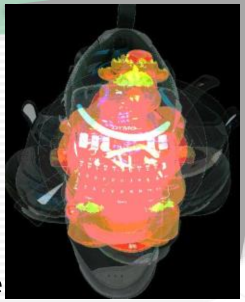


Figure 7: “Morph” template

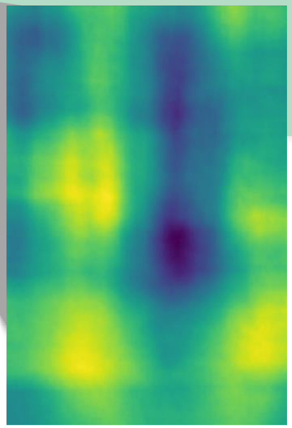


Figure 8: Template Match Result

Figure 10: Located Image, 3D background blurred

RESULTS

OBJECT RECOGNITION: CURE-OR

CREATING THE NEXT®

RESULTS

- So, does it work?
  - **No**
    - 22% prediction accuracy with model alone, 24% with full processing pipeline
    - Not enough 3D background is removed
    - If background has ‘perfect trim’ (Fig. 11) identification works; current background smoothing (Fig. 12) is not enough
    - **NOTE:** Model was never trained on validation data prior to testing - missed potential for slightly better results
- Does anything work?
  - **Yes**
    - Accuracy on *validation set* goes from 80% to 90% when distortion prediction + image cleanup is used
      - Negligible affect on test set prediction (3D backgrounds too much to overcome with current model)

Figure 11: Shoe

Figure 12: Label Maker

RESULTS – CONT.

Table 1: Prediction Misses on Validation Set (175 misses, 82% accurate)

Distortion	None	Resize	UnderSat	OverSat	Gauss	Contrast	LensDirt1	LensDirt2	S&P
Misses	0	54	0	8	76	2	2	1	32
Percent	0	31%	0%	5%	43%	1%	1%	1%	18%

Table 2: Prediction Misses on Validation Set after distortion prediction and image cleaning (114 misses, 89% accurate)

Distortion	None	Resize	UnderSat	OverSat	Gauss	Contrast	LensDirt1	LensDirt2	S&P
Misses	0	2	0	8	76	3	6	1	18
Percent	0%	2%	0%	7%	67%	3%	5%	1%	16%

Georgia Tech

LESSONS LEARNED / ATTEMPTS

“HOW NOT TO CLASSIFY A CAT”

OBJECT RECOGNITION: CURE-OR

CREATING THE NEXT®

LESSONS LEARNED/ ATTEMPTED METHODS

- Template matching method requires way to much tuning to generalize well
- Many algorithms tried, failed, and were tried again
  - Noise added in an attempt to mask 3D background (Fig 13)
  - Match-Template outline used for mask, to minimize background (Fig 14)
  - Minimize background by tracing outline of object based on coordinate given from template matching
    - Sobel filters used to try and separate object from same-intensity 3D backgrounds, e.g. the black book shelf (Fig 15), to then try and trace the outline of the object
    - Intensity quantization (Fig 16)
      - Intent was to merge similar intensities (ideally the entire object in question, and background) such that they could be separated
  - Individual object templates instead of massive morph (Fig 17)

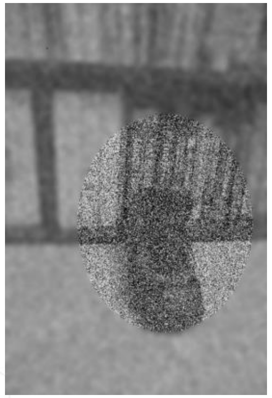


Figure 13: Dithered, masked, and smoothed test image

Figure 14: Test image masked with 'match template' outline

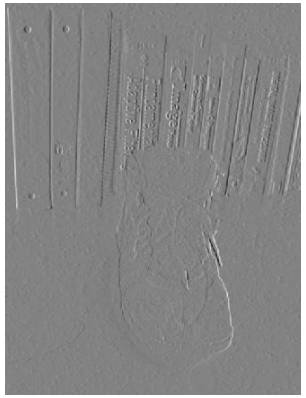


Figure 15: Sobel-x of test image. Note the removal of horizontal bookcase lines from background

Figure 16: Quantized test image.

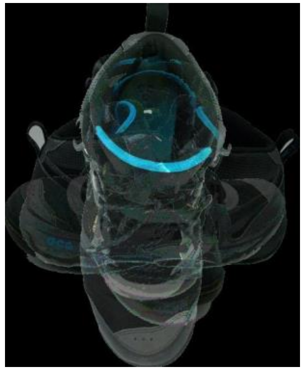


Figure 17: Object template example

CONCLUSION / NEXT STEPS

OBJECT RECOGNITION: CURE-OR

CONCLUSION

- In absence of robust training set, identification accuracy depends solely on ability to remove (or nullify) conditions not present during training, i.e. 3D background
- Distortions can be overcome with correct pre-prediction processing, but if objects are in environment not trained for, accuracy will be low
- Majority of the algorithm that I thought was novel already existed:
  - Blob detection / template matching with correlation
  - Morphed template
  - Background reduction/smoothing

NEXT STEPS

- If I had to start over (with a beefier computer)…
  - Focus on how the models are trained / image augmentation
    - Train model with all images distorted / dithered such that dithering in the test set (to suppress background) could be used to greater affect
    - Generate much larger training sets by merging objects with varying 3D backgrounds and background 'distractions'
  - Train with deeper and wider models, train for much longer
- If I had to keep going with the template matching + background smoothing idea…
  - Pick the 'object outline' tracing idea back up, such that a much tighter bounding rectangle or mask could be placed around the located image (similar to Fig 11)
  - Template match with outlines (dark spaces are consistently confused with shoes)
  - Train with downsampled images to train faster/longer and get a better model

QUESTIONS?

EMAIL: JLEIKER3@GATECH.EDU

GTID : 903453031

CREATING THE NEXT®