
# Generative Sensing: Reliable Recognition from Unreliable Sensor Data

Arizona State University

Lina Karam

May 2018

# Sensors and Data … Everywhere

- Accessibility of compact and low-cost sensors enabling sensing on the go.
  - Cameras and IR sensors on portable devices
  - Embedded low-power mm-wave radar for natural-human interactions (e.g., Google ATAP Soli)
- Advances in machine learning driving more data to be “sensed” for various applications
  - Self-Driving Cars
  - Assistive Technologies
  - Security, Authentication, Surveillance
  - Natural Human-Machine Interfaces

# Various Types of Sensors

- Modality
  - Visible Spectrum Camera
  - Infra Red (IR)
  - Radar
  - Lidar
  - Ultrasound
- Cost, Size, SNR, Energy Efficiency
  - High-end: high resolution, high SNR but high cost, large size, and high power consumption
  - Low-end: low cost, low profile (thin, small size), low power but low resolution (blur), low SNR (noise)

# Exponential Growth of Unconstrained Consumer-Grade Data


"It would take an individual over 5 million years to watch the amount of video that will cross global IP networks each month in 2019"

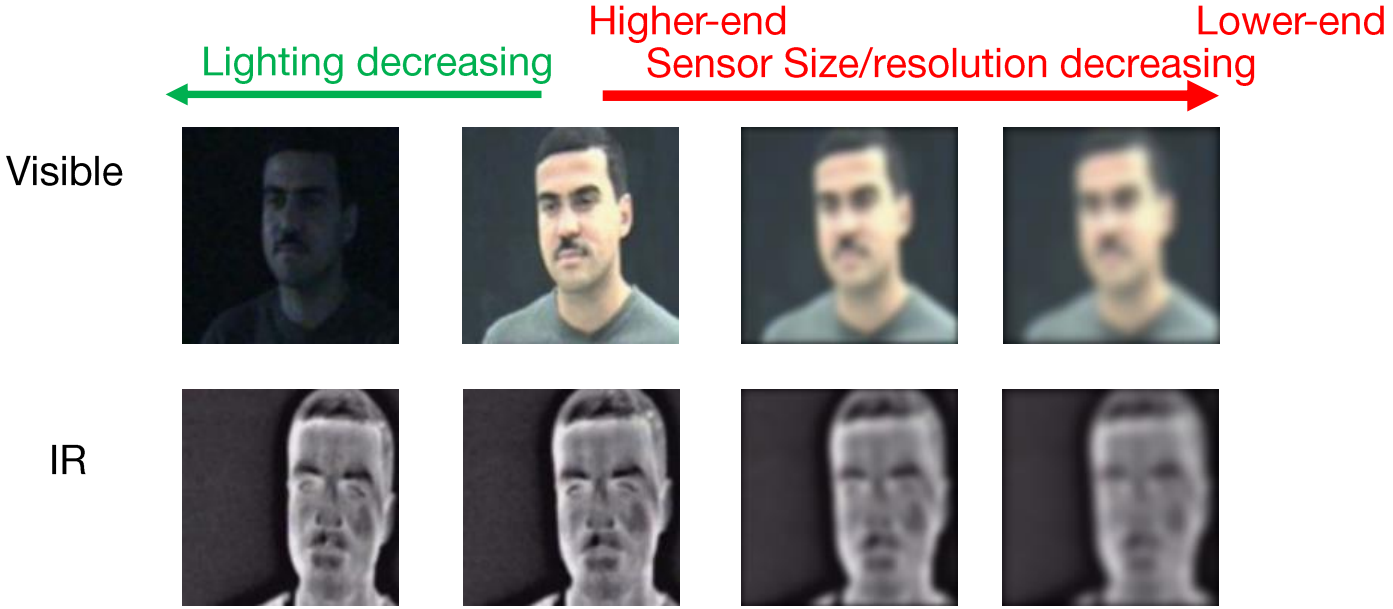
*ernet*
*l be 80*
*onsumer*
*in 2019"*

**Source:**
Cisco VN Mobile Forecast, White Paper, May 2015

**Mostly "Consumer-Grade"** Videos in the Wild !

# Various Types of Sensors

- Data quality impacted by environment and sensor specifications




- Multi-modal sensors needed for robustness under varying environmental conditions (e.g., variations in lighting, weather,…), which can be costly.

# What about Machine Learning ?

Can state-of-the-art deep learning neural networks (DNNs) deal with variations in sensor types, environment conditions, and, consequently, data quality?

# How Robust are DNNs?

- Deep Neural Networks (DNNs) sensitive even to small amounts of blur (sensor size, low-resolution) and noise (low SNR)

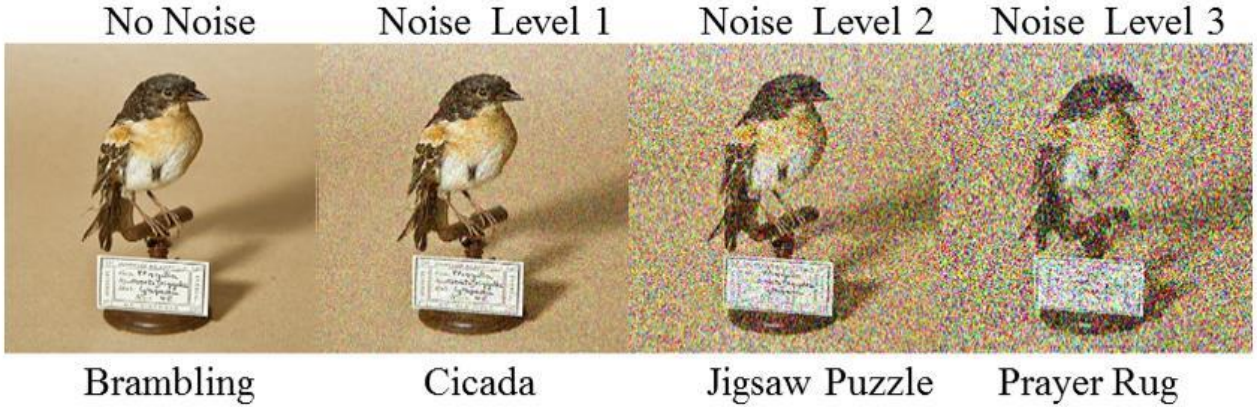

DNN Prediction using AlexNet trained on ImageNet.

S. Dodge and L. Karam. “Understanding how image quality affects deep neural networks.” *QoMEX,* 2016.

T. Borkar and L. Karam, “DeepCorrect: Correcting DNN Models against Image Distortions,” arXiv:1705.02406, May 2017.

# How Robust are DNNs?

- Deep Neural Networks (DNNs) very sensitive even to small amounts of blur (small sensor size, low-resolution) and noise (low SNR)

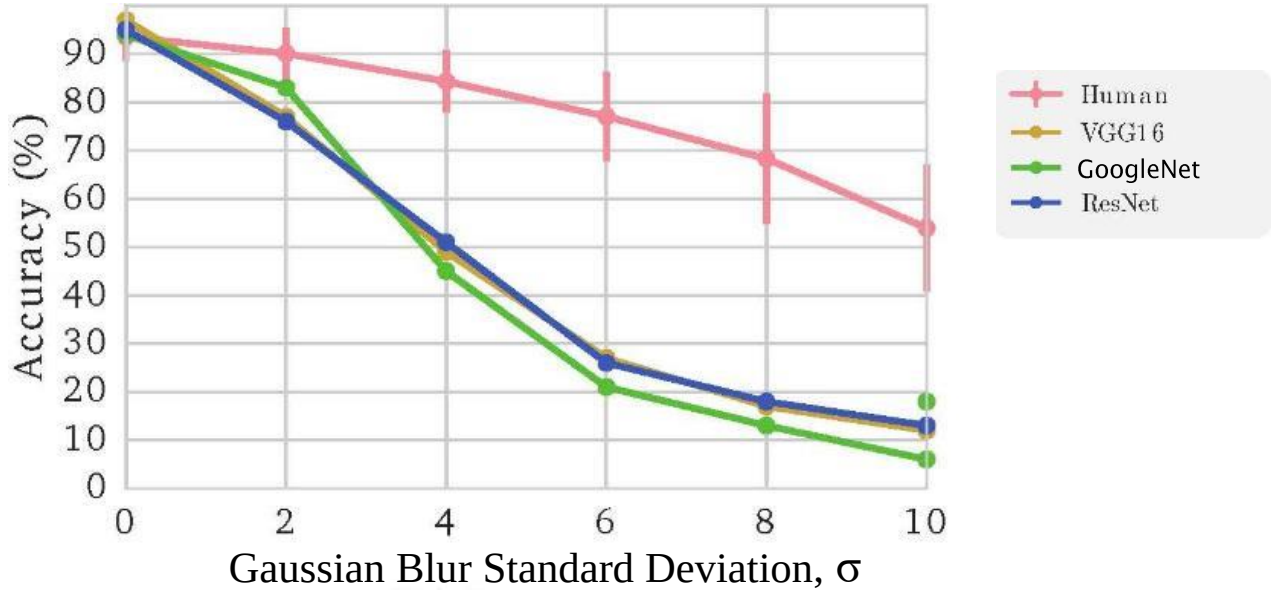



DNN Prediction using AlexNet trained on ImageNet.

S. Dodge and L. Karam. "Understanding how image quality affects deep neural networks." *QoMEX,* 2016.

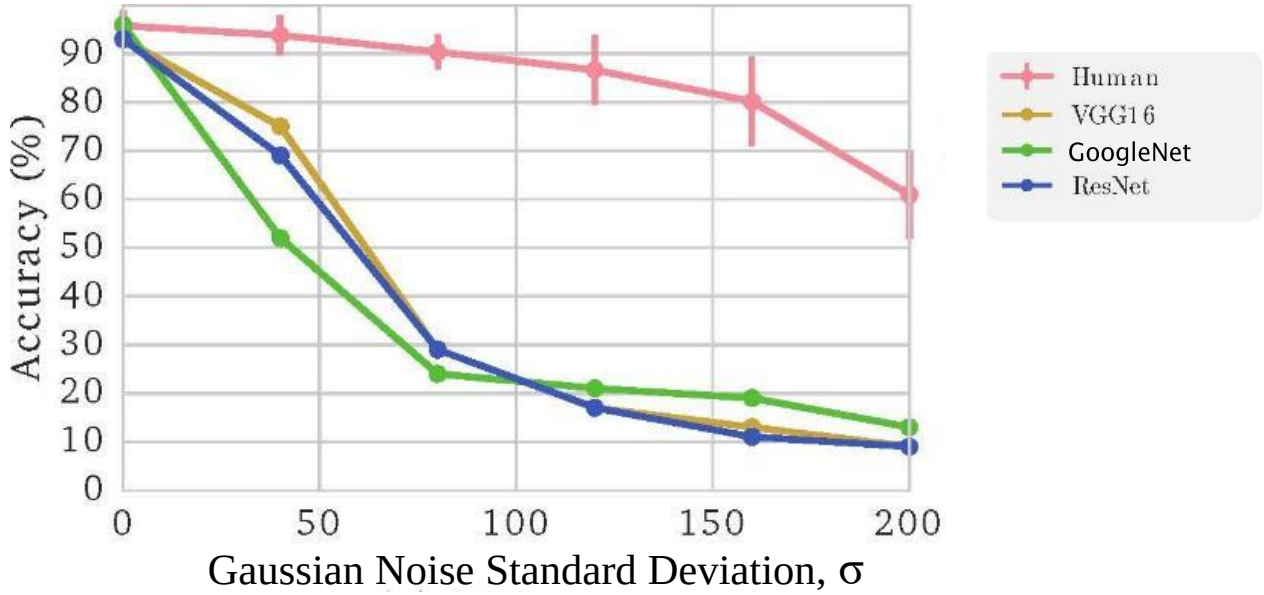
T. Borkar and L. Karam, "DeepCorrect: Correcting DNN Models against Image Distortions," arXiv:1705.02406, May 2017.

# Effect of Blur (Sensor Size/Resolution) on Human and DNN Prediction



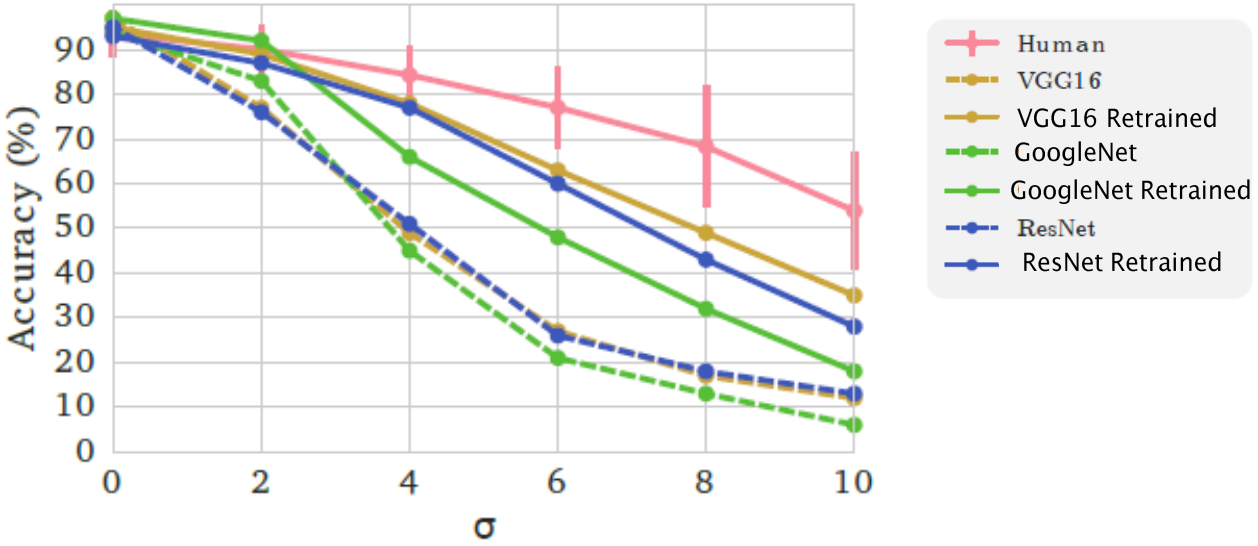

S. Dodge & L. Karam. “A Study and Comparison of Human and Deep Learning Recognition Performance Under Visual Distortions,” *ICCCN,* 2017.
K. Simonyan and A. Zisserman. “Very Deep Convolutional Networks for Large-Scale Image Recognition.” arXiv, 2014. (VGG 16)
C. Szegedy et al. “Going deeper with convolutions.” IEEE CVPR, 2016. (GoogleNet)
Kaiming et al. “Deep residual learning for image recognition.” IEEE CVPR, 2016. (ResNet 50)

# Effect of Noise (Sensor SNR) on Human and DNN Prediction



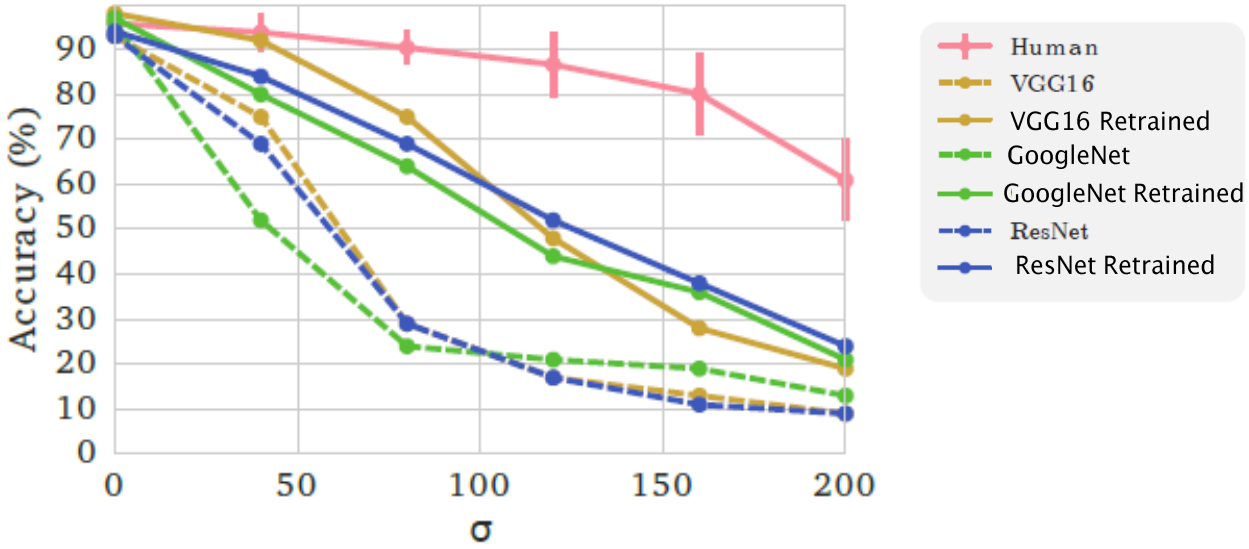

S. Dodge & L. Karam. "A Study and Comparison of Human and Deep Learning Recognition Performance Under Visual Distortions," *ICCCN,* 2017.
K. Simonyan and A. Zisserman. "Very Deep Convolutional Networks for Large-Scale Image Recognition." arXiv, 2014. (VGG 16)
C. Szegedy et al. "Going deeper with convolutions." IEEE CVPR, 2016. (GoogleNet)
Kaiming et al. "Deep residual learning for image recognition." IEEE CVPR, 2016. (ResNet 50)

# Effect of Blur (Sensor Size/Resolution) on Human and DNN Prediction




S. Dodge and L. Karam. "A Study and Comparison of Human and Deep Learning Recognition Performance Under Visual Distortions," International Conference on Computer Communications and Networks *(ICCCN),* 2017.

# Effect of Noise (Sensor SNR) on Human and DNN Prediction




S. Dodge and L. Karam. "A Study and Comparison of Human and Deep Learning Recognition Performance Under Visual Distortions," International Conference on Computer Communications and Networks *(ICCCN),* 2017.

# Human vs DNN Performance

- Human subjects can recognize distorted images even when the display time is restricted and is very short (100ms), so in this case the human subject is using “gist” level information.
- Deep neural networks perform worse than humans on distorted images for both free-viewing and 100ms “gist” viewing.
- Training the networks on distorted images helps, but there is still a gap in performance especially when classes are not easy to separate.

S. Dodge and L. Karam. “Can the Early Human Visual System Compete with Deep Neural Networks?” International Conference on Computer Vision (ICCV), Workshop on Mutual Benefits of Cognitive and Computer Vision (MBCC), Oct. 2017. Available at: http://openaccess.thecvf.com/ICCV2017_workshops/ICCV2017_W40.py.

# From Restoring Data to Restoring DNN Features

- DNN feature maps also called DNN filter activations (output of filters in each convolutional layer)
- Questions to consider:
  - Are networks able to learn filters whose activations are invariant to distortions (blur or noise) in the input data even when such distortions are absent from training set?
  - Do distortions affect significantly some features more than others?
- Can we identify the convolutional filters that are most susceptible to image distortions and recover the lost performance, by only correcting the outputs (i.e., feature maps, activations) of such filters?
- Can we transform low-quality features into high-quality features ?
  - quality measured in terms of classification performance

T. Borkar and L. Karam, "DeepCorrect: Correcting DNN Models against Image Distortions," arXiv:1705.02406, May 2017.

# DeepCorrect: Selectively Correcting Most Susceptible DNN Features

- Convolutional filters that are most sensitive to degradations in image quality can be identified based on gain in classification accuracy when features are restored.
- Correcting only a fraction of the most susceptible filter activations results in a significant performance improvement on popular datasets including ImageNet.
- DeepCorrect model can achieve a classification accuracy higher than fine-tuning/retraining (which retrains all network parameters).
- In a lot of cases, not enough data is available for retraining the full network; proposed framework alleviates this issue.

T. Borkar and L. Karam, "DeepCorrect: Correcting DNN Models against Image Distortions," arXiv:1705.02406, May 2017.

# Generative Sensing Framework

- The Generative Sensing framework generalizes DeepCorrect to sensors under varying conditions (e.g., varying illumination, environment, and acquisition characteristics in addition to resolution) and varying modalities (e.g., visible, NIR, IR, Ultrasound, RADAR; speech/audio).

T. Borkar and L. Karam, "DeepCorrect: Correcting DNN Models against Image Distortions," arXiv:1705.02406, May 2017.

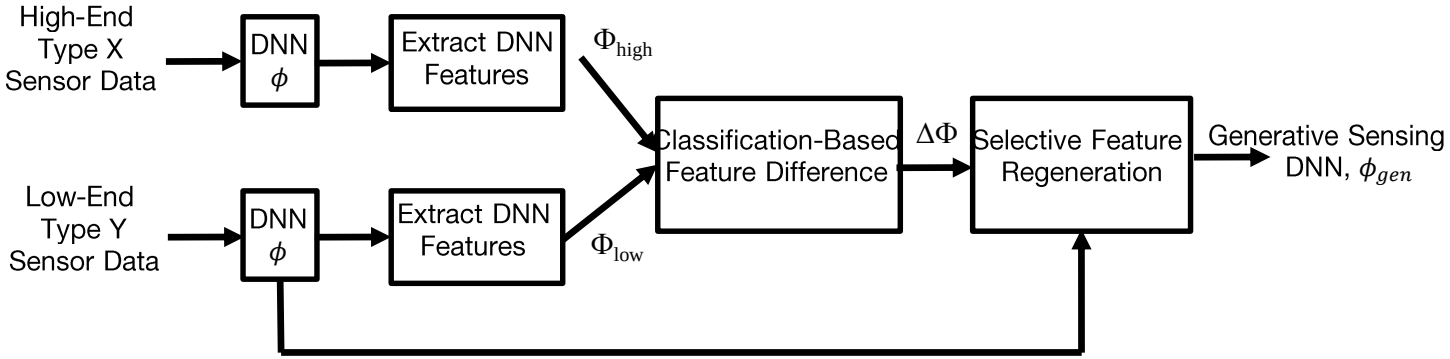
# Generative Sensing Framework

- Goal: Transform  low-end, low-quality sensor/data into high-end, high-quality sensor/data of the same or different modality in order to attain a recognition accuracy close to that of the high-end sensor/data.
- Strategy:
  - Selective regeneration of DNN features of low-end sensor/data to match those of high-end sensor/data in terms of increased classification accuracy
  - The aim is NOT to generate the high-end sensor data but rather the corresponding high-quality DNN features that most significantly affect the classification accuracy.

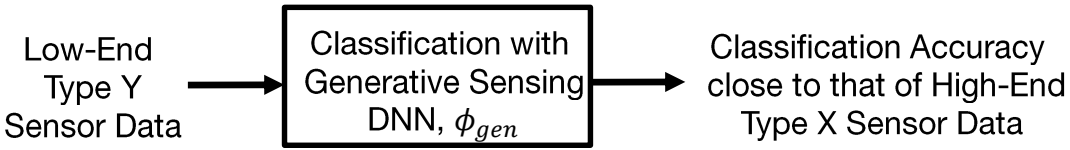
# Generative Sensing Framework

- Existing methods focus on data generation and maximize a similarity measure between reference data and generated data
  - Colorization: grayscale to color (Deshpande et al, ICCV 2015; Zhang et al, ECCV 2016), IR colorization (Limmer & Lensch, ICMLA 2016)
  - Artistic Style Transfer (Gatys et al., CVPR 2016)
  - Data Augmentation: e.g., thermal image generation (Kniaza et al., PSBB, 2017)
- Existing classifier-friendly pre-processing methods (Diamond et al., arXiv, 2017)
- Proposed Generative Sensing
  - DNN-based feature generation rather than data generation
  - maximize the classification accuracy rather than a similarity measure
  - based on discriminative models and optimizes a target-oriented objective function (e.g., a regularized categorical cross-entropy loss function)

# Generative Sensing Framework



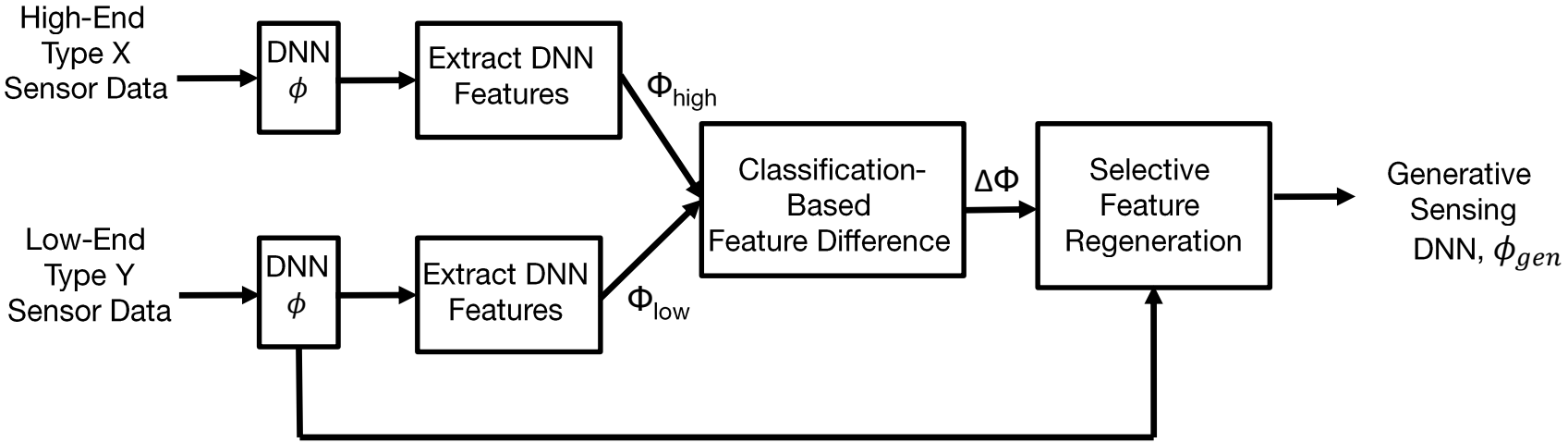

Generative Sensing – Training Phase
(DNN $\phi$ pre-trained using Type X data)




Generative Sensing – Testing Phase/Deployment

L. Karam et al., "Generative Sensing: Transforming Unreliable Sensor Data for Reliable Recognition," IEEE MIPR, 2018.

# Generative Sensing Framework

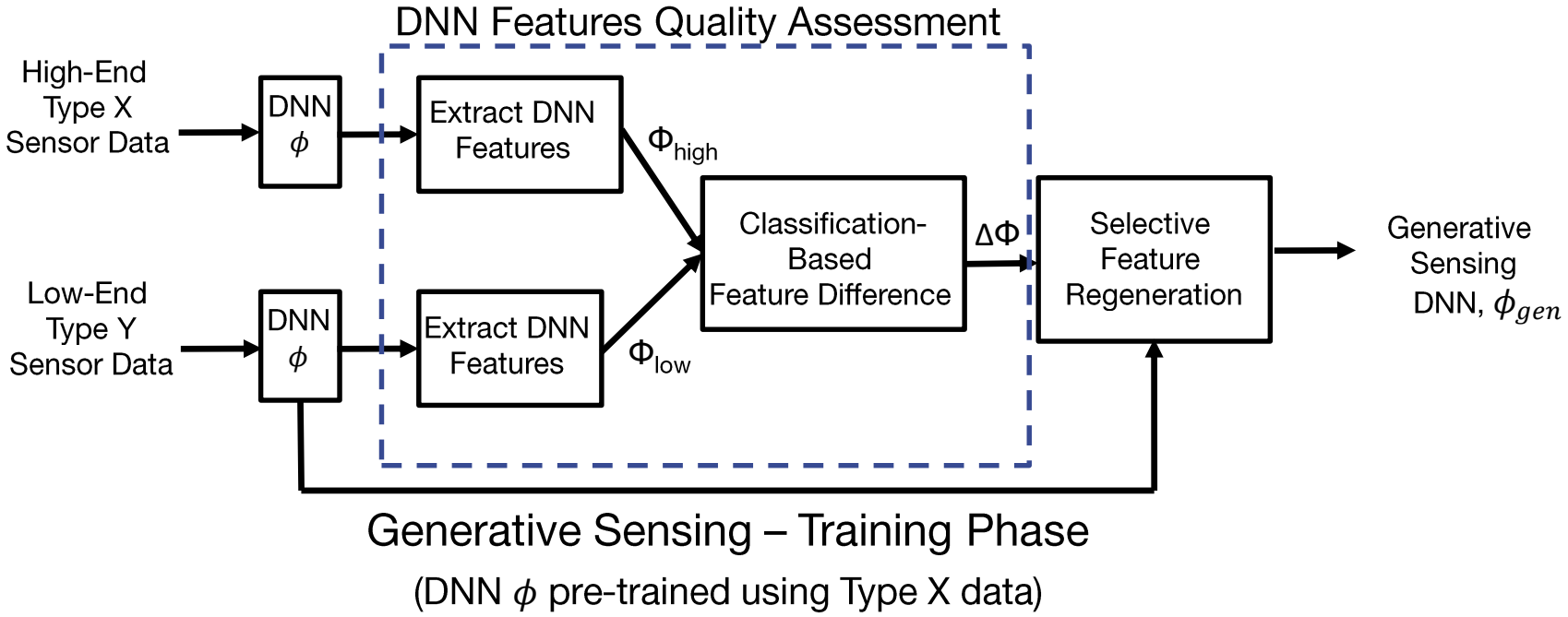



Generative Sensing – Training Phase

(DNN $\phi$ pre-trained using Type X data)

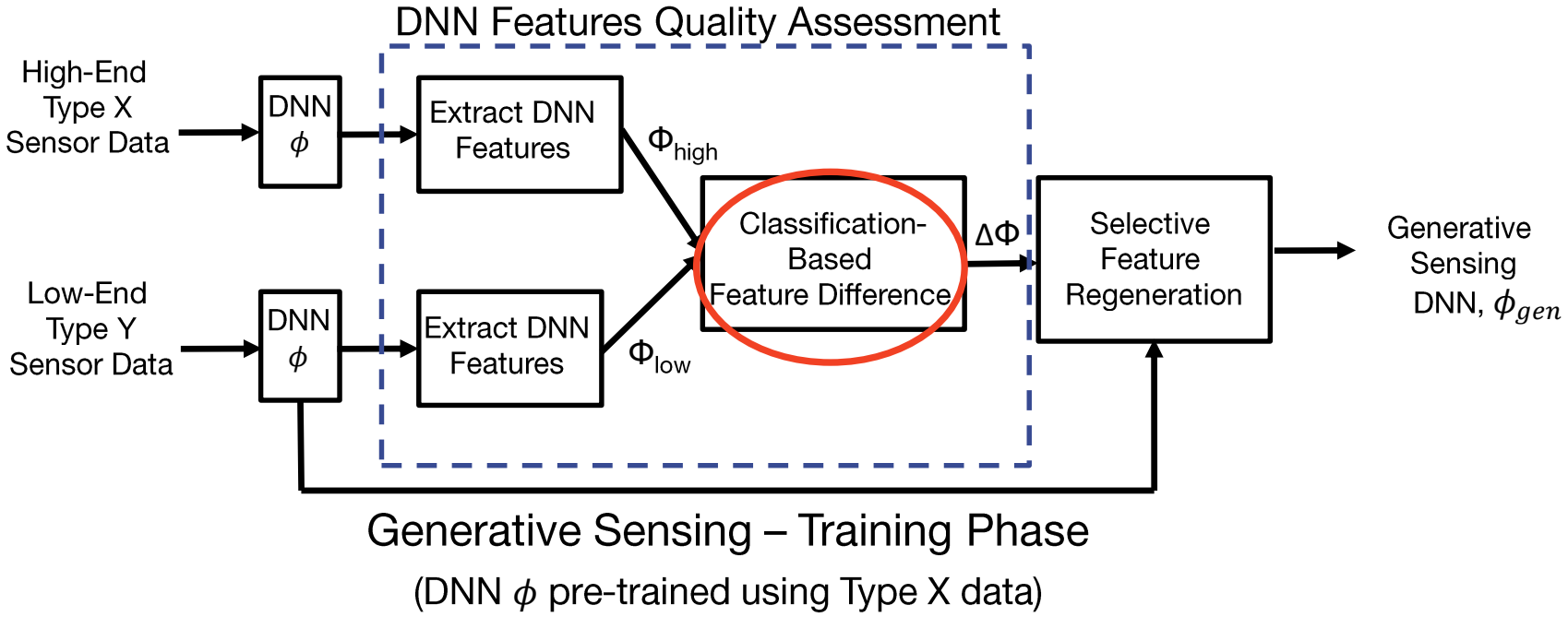
L. Karam et al., "Generative Sensing: Transforming Unreliable Sensor Data for Reliable Recognition," IEEE MIPR, 2018.

# Generative Sensing Framework



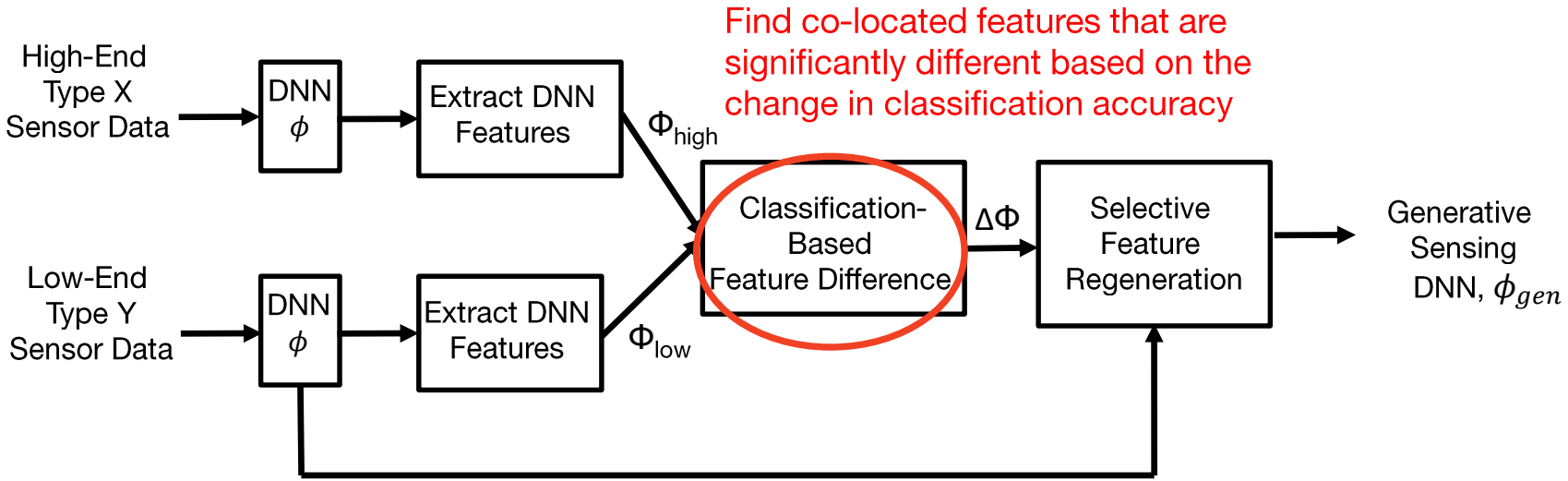

L. Karam et al., "Generative Sensing: Transforming Unreliable Sensor Data for Reliable Recognition," IEEE MIPR, 2018.

# Generative Sensing Framework




L. Karam et al., "Generative Sensing: Transforming Unreliable Sensor Data for Reliable Recognition," IEEE MIPR, 2018.

# Generative Sensing Framework

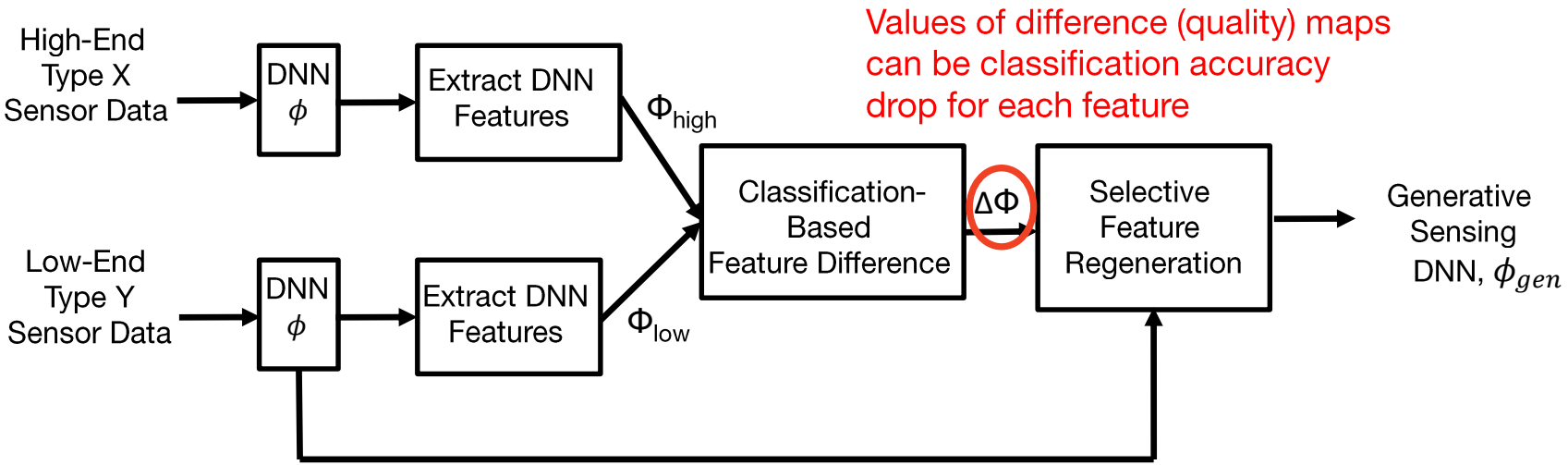



Generative Sensing – Training Phase

(DNN $\phi$ pre-trained using Type X data)

L. Karam et al., "Generative Sensing: Transforming Unreliable Sensor Data for Reliable Recognition," IEEE MIPR, 2018.

# Generative Sensing Framework

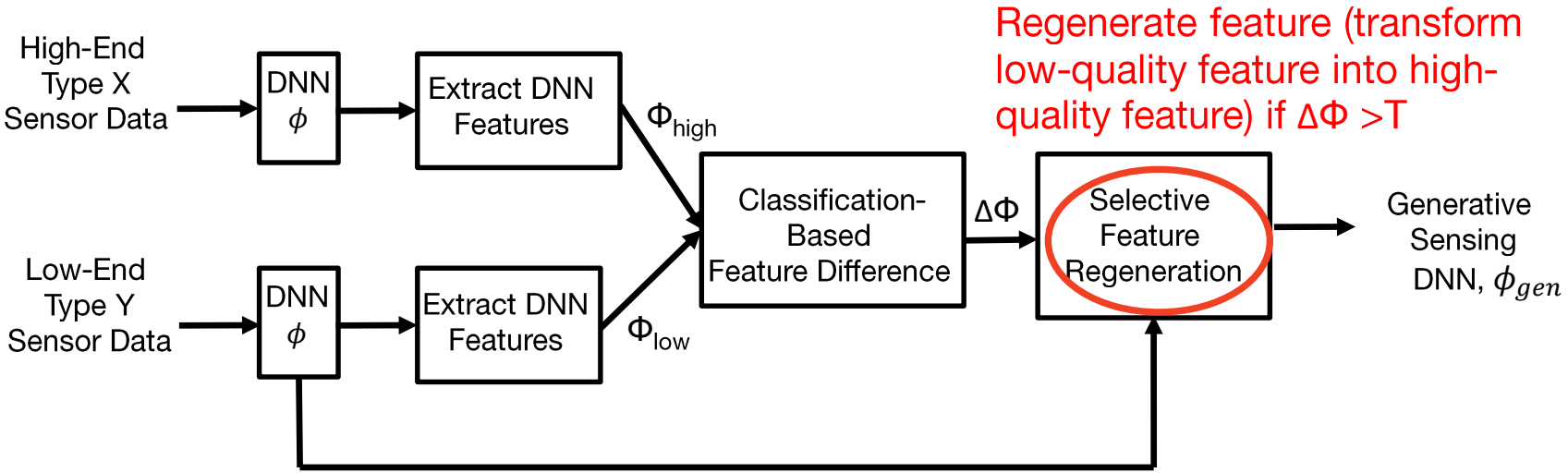



Generative Sensing – Training Phase

(DNN $\phi$ pre-trained using Type X data)

L. Karam et al., "Generative Sensing: Transforming Unreliable Sensor Data for Reliable Recognition," IEEE MIPR, 2018.

# Generative Sensing Framework

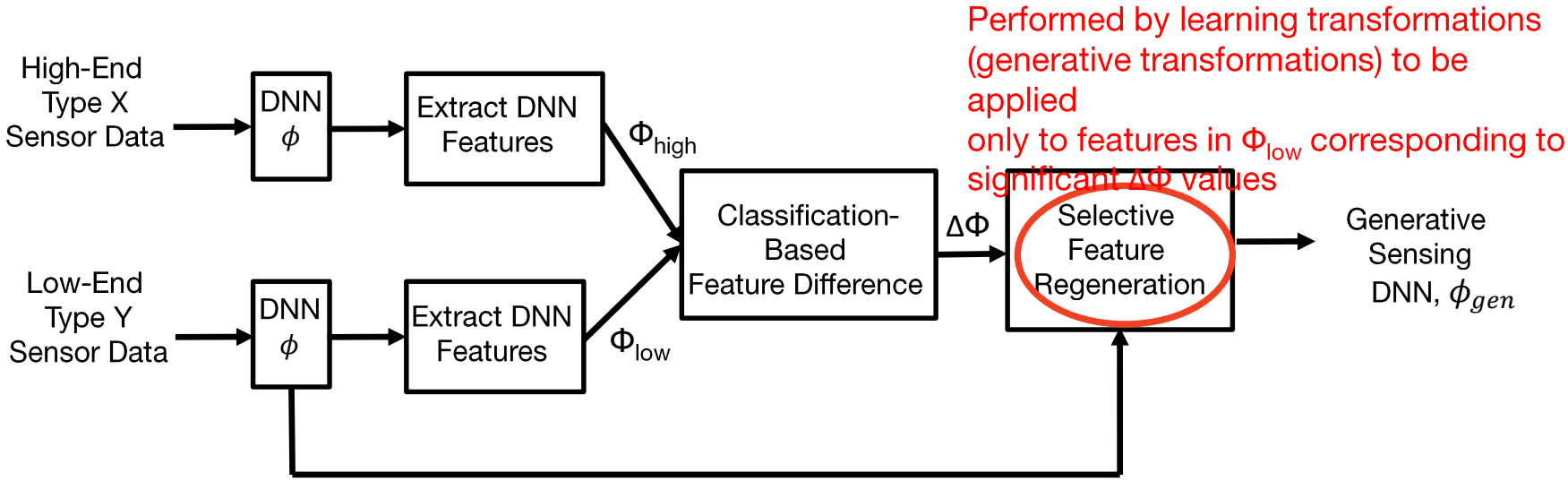



Generative Sensing – Training Phase

(DNN $\phi$ pre-trained using Type X data)

L. Karam et al., "Generative Sensing: Transforming Unreliable Sensor Data for Reliable Recognition," IEEE MIPR, 2018.

# Generative Sensing Framework

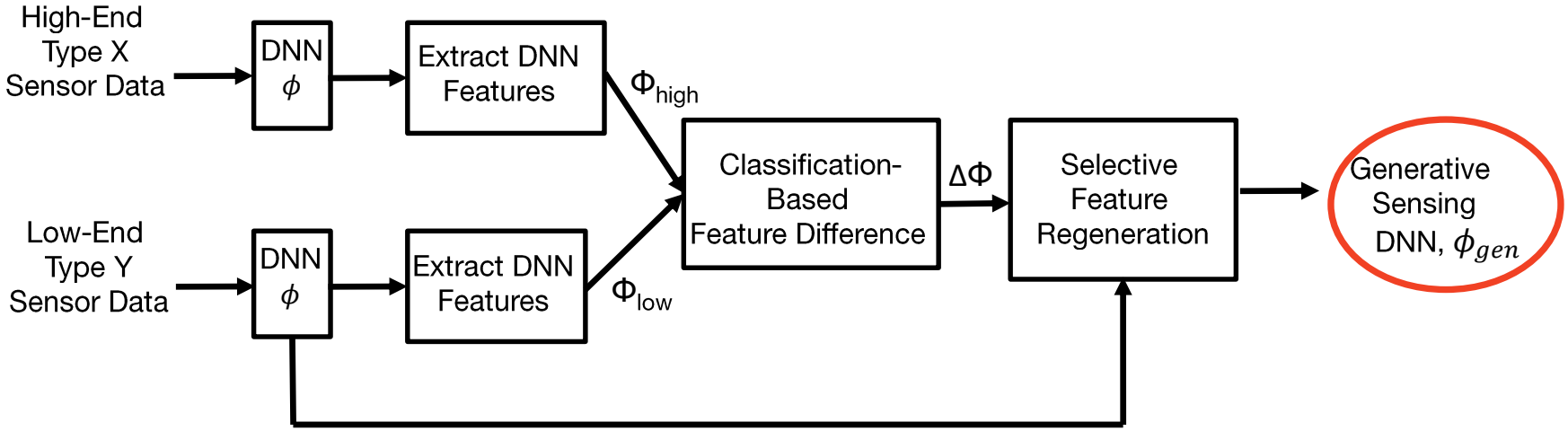



Generative Sensing – Training Phase

(DNN $\phi$ pre-trained using Type X data)

L. Karam et al., "Generative Sensing: Transforming Unreliable Sensor Data for Reliable Recognition," IEEE MIPR, 2018.

# Generative Sensing Framework

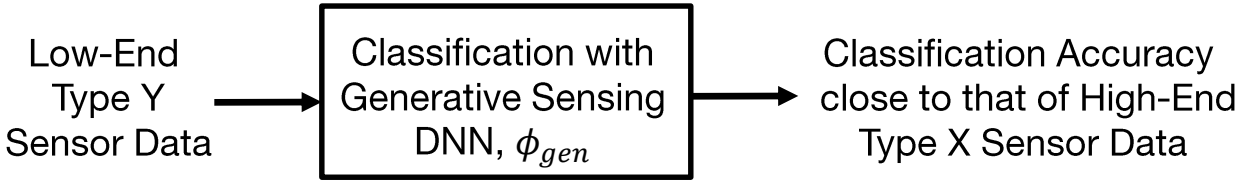



Generative Sensing – Training Phase

(DNN $\phi$ pre-trained using Type X data)

L. Karam et al., "Generative Sensing: Transforming Unreliable Sensor Data for Reliable Recognition," IEEE MIPR, 2018.

# Generative Sensing Framework




Generative Sensing – Testing Phase/Deployment

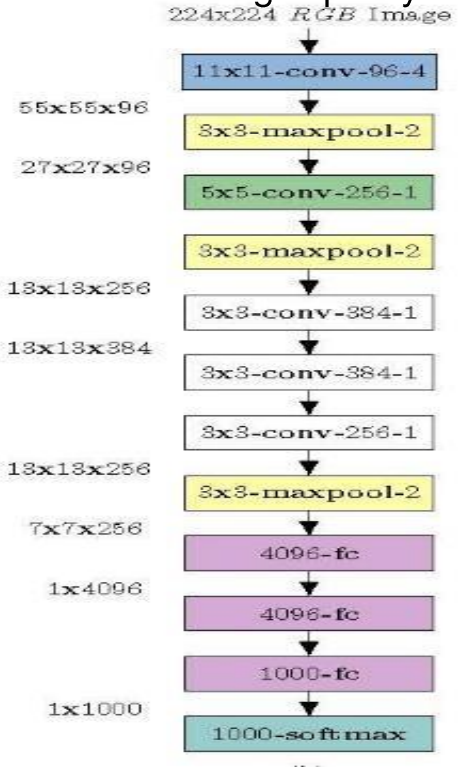
L. Karam et al., "Generative Sensing: Transforming Unreliable Sensor Data for Reliable Recognition," IEEE MIPR, 2018.

# Generative Sensing Framework

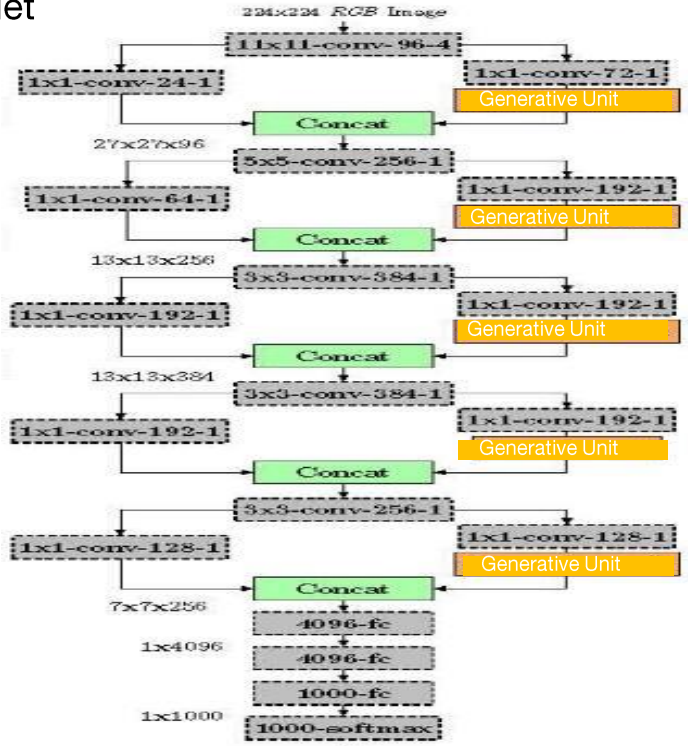
- Selective Feature Regeneration
  - Performed by learning transformations to be applied only to select features in $\Phi_{low}$ (DNN features resulting from low-end sensor data) that are associated with significantly large $\Delta\phi$ values while leaving all other features unchanged.
  - Example: Transformations can be learned by using small convolutional or residual learning blocks, which we refer to as generative units.
  - Generative Sensing Network $\phi_{gen}$ = original pre-trained DNN $\phi$ + learned generative units applied only to select feature maps

# Generative Sensing with AlexNet

Baseline: Original AlexNet
Pre-trained on high-quality ImageNet

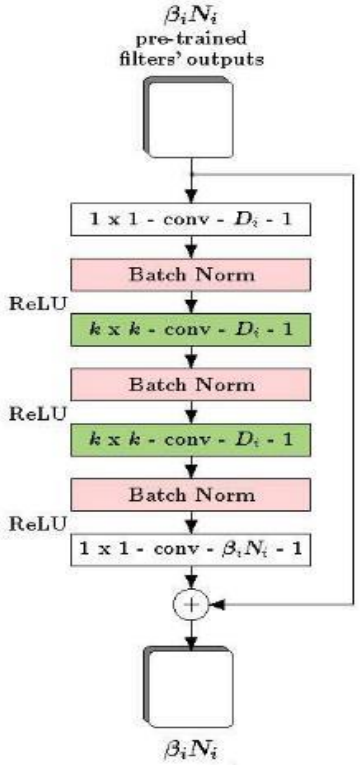



AlexNet with Generative Sensing



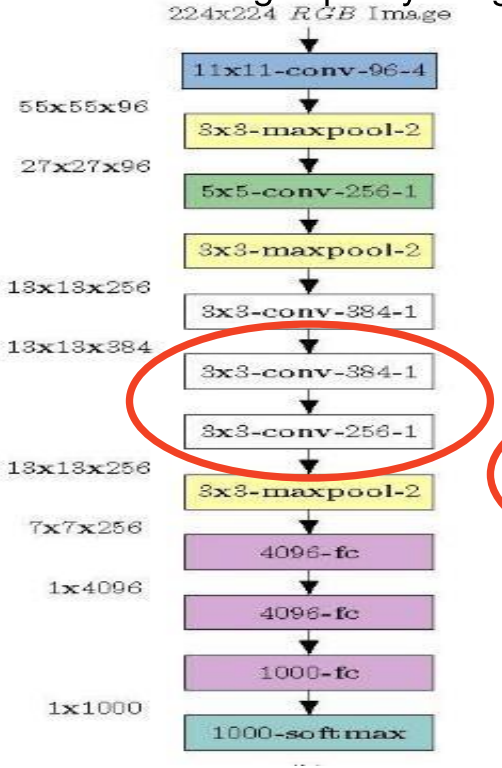

Only Generative Units trained on low-quality data

**Generative Unit**

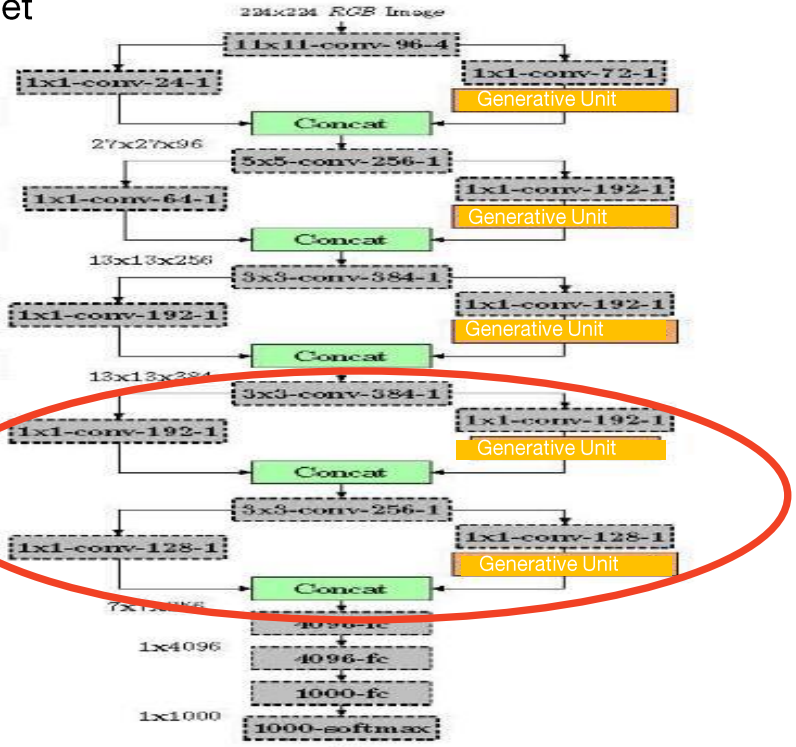


# Generative Sensing with AlexNet

Baseline: Original AlexNet
Pre-trained on high-quality ImageNet

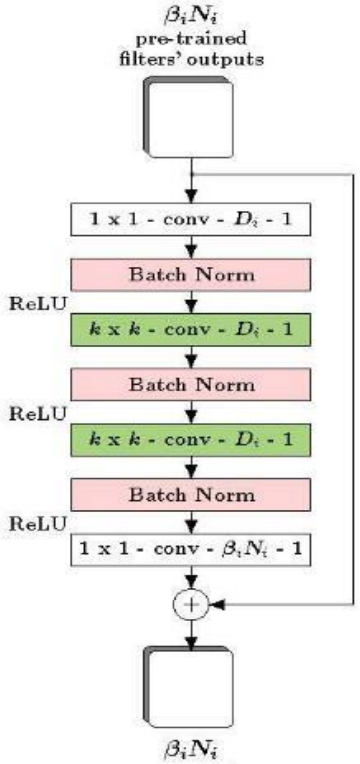



AlexNet with Generative Sensing



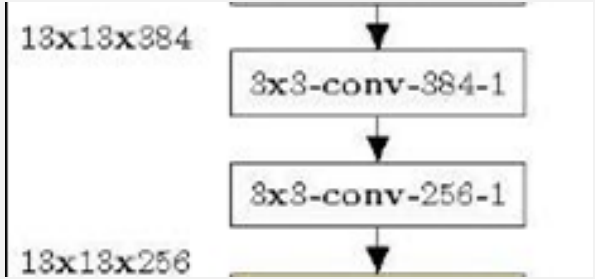

Only Generative Units trained on low-quality data

**Generative Unit**



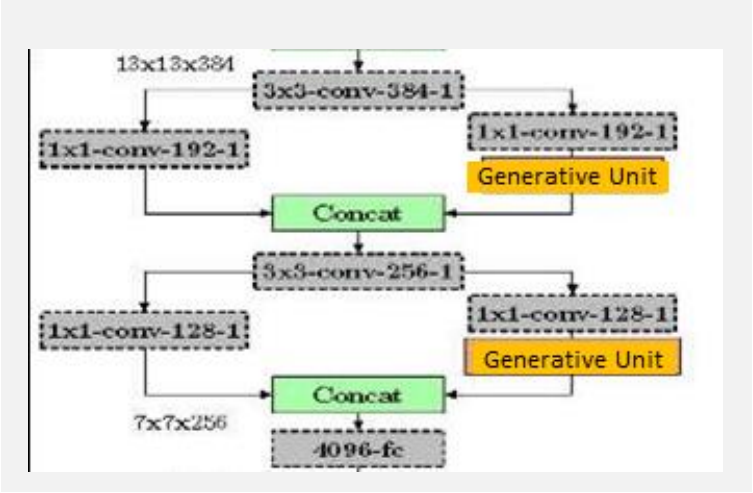
# Generative Sensing with AlexNet

Baseline: Original AlexNet



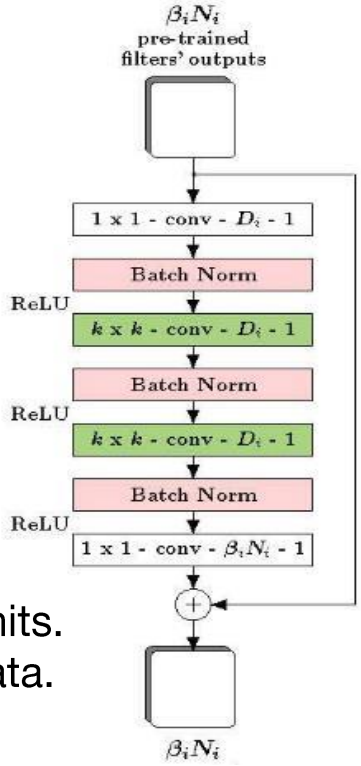

Pre-trained on RGB high-quality ImageNet

AlexNet with Generative Sensing




Original Pre-trained AlexNet plus Generative Units. Only Generative Units trained on low-quality data.

**Generative Unit**



# Generative Sensing Framework

- Learning Generative Units Parameters
  - The transformation can be estimated by determining the trainable parameters $\mathbf{W}_{gen}$ of each generative unit so as to minimize a target-oriented loss function:

$$E(\mathbf{W}_{gen}) = \lambda\, \rho(\mathbf{W}_{gen}) + \frac{1}{M}\sum_{i=1}^{M} \mathcal{L}(\, \phi_{gen}(\mathbf{x}_i), y_i)$$

$\mathcal{L}(.,.)$: classification loss function

$y_i$: target output label for input $\mathbf{x}_i$

$\phi_{gen}(.)$: output of the network with the selectively applied generative units, which we refer to as generative sensing network

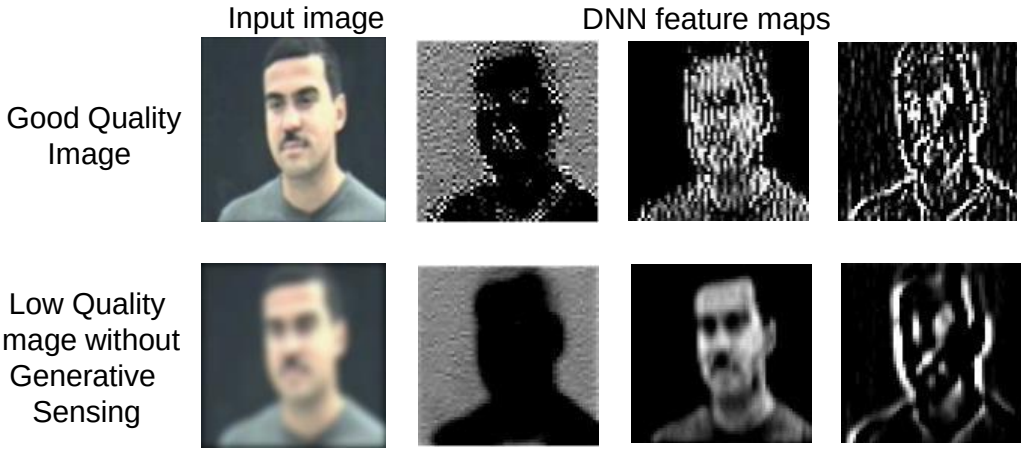
$\rho(.)$: regularization term (e.g., $\ell 1$ or $\ell 2$ norm)

$\lambda$: regularization parameter

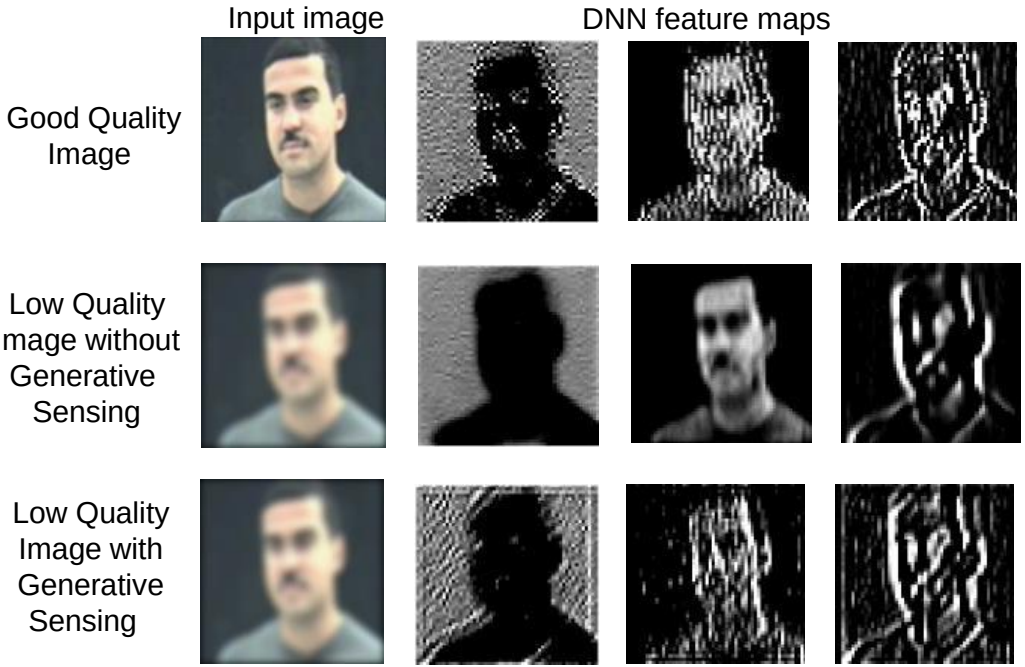
$M$: total number of data samples in the training set.

  - The generative units designed to have small number of parameters compared to DNN $\phi$.

# Effect of Distortions on DNN Feature Maps



# Feature Regeneration based on Generative Sensing Framework



# Top-1 Accuracy Results for ImageNet (ILSVRC2012) Object Recognition under Blur

| Method | Blur Distortion Level | | | | | | | Avg | # Params |
|---|---|---|---|---|---|---|---|---|---|
| | $\sigma_b = 0$ | $\sigma_b = 1$ | $\sigma_b = 2$ | $\sigma_b = 3$ | $\sigma_b = 4$ | $\sigma_b = 5$ | $\sigma_b = 6$ | | (million) |
| Baseline | 0.5694 | 0.4456 | 0.2934 | 0.1585 | 0.0786 | 0.0427 | 0.0256 | 0.2305 | 60.96 |
| **Fully Retrained** | 0.5553 | 0.5301 | 0.5004 | 0.4669 | 0.4276 | 0.3886 | 0.3485 | 0.4596 | 60.96 |
| **Gen. Sense Net** | **0.5724** | **0.5522** | **0.5240** | **0.5100** | **0.4937** | **0.4643** | **0.4334** | **0.5071** | **2.81** |

L. Karam et al., "Generative Sensing: Transforming Unreliable Sensor Data for Reliable Recognition," IEEE MIPR, 2018.

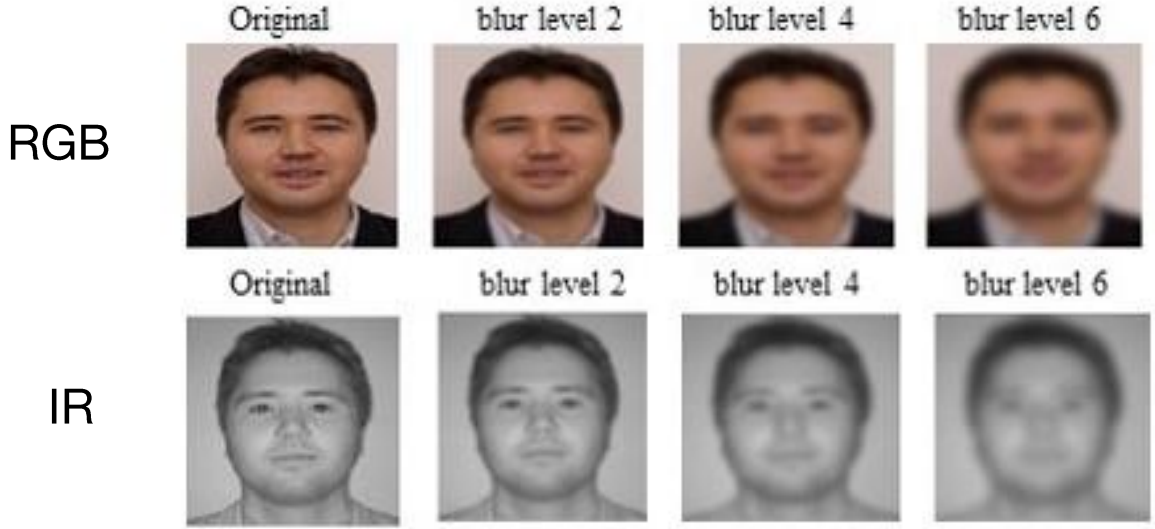
# Top-1 Accuracy Results for ImageNet (ILSVRC2012) Object Recognition under Noise

| Method | Noise Distortion Level | | | | | | | Avg | # Params |
|---|---|---|---|---|---|---|---|---|---|
| | $\sigma_n = 0$ | $\sigma_n = 10$ | $\sigma_n = 20$ | $\sigma_n = 40$ | $\sigma_n = 60$ | $\sigma_n = 80$ | $\sigma_n = 100$ | | (million) |
| Baseline | 0.5694 | 0.5218 | 0.3742 | 0.1256 | 0.0438 | 0.0190 | 0.0090 | 0.2375 | 60.96 |
| **Fully Retrained** | 0.5540 | 0.5477 | 0.5345 | 0.5012 | 0.4654 | 0.4297 | 0.3936 | 0.4894 | 60.96 |
| **Gen. Sense Net** | 0.5712 | 0.5660 | 0.5546 | **0.5213** | **0.4870** | **0.4509** | **0.4138** | **0.5092** | **2.81** |

L. Karam et al., "Generative Sensing: Transforming Unreliable Sensor Data for Reliable Recognition," IEEE MIPR, 2018.

# Face Recognition using the SCface Dataset.

- The SCface dataset consists of face images acquired in the visible (RGB) as well as infrared spectrum (IR) for 130 subjects. We make use of the frontal mugshot images.



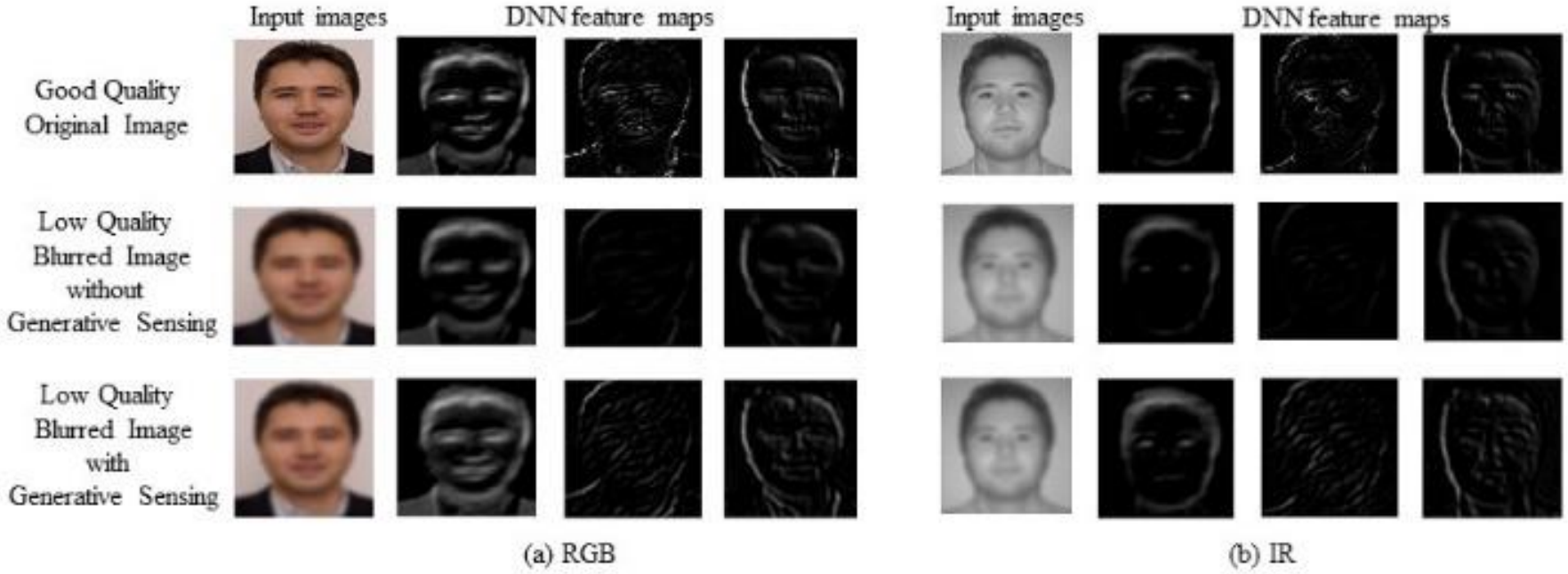

M. Grgic, K. Delac, S. Grgic, SCface - surveillance cameras face database, Multimedia Tools and Applications Journal, vol. 51, No. 3, pp. 863-879, February 2011.

# Top-1 Accuracy Results for Face Recognition using the SCface Dataset.

| Method-Modality | Sensor Resolution: Blur Level | | | | | | | Avg |
|---|---|---|---|---|---|---|---|---|
| | $\sigma_b = 0$ | $\sigma_b = 1$ | $\sigma_b = 2$ | $\sigma_b = 3$ | $\sigma_b = 4$ | $\sigma_b = 5$ | $\sigma_b = 6$ | |
| Baseline-RGB | **0.9923** | 0.7538 | 0.4384 | 0.3230 | 0.1461 | 0.1000 | 0.0770 | 0.4043 |
| *Generative Sensing*- RGB | 0.9538 | **0.9461** | **0.9000** | **0.8692** | **0.7692** | **0.6846** | **0.6461** | **0.8241** |
| Baseline-IR | **0.9769** | 0.7923 | 0.4769 | 0.1076 | 0.0461 | 0.0076 | 0.0076 | 0.3450 |
| *Generative Sensing*-IR | 0.9000 | **0.8777** | **0.8077** | **0.7538** | **0.6538** | **0.5077** | **0.4692** | **0.7098** |

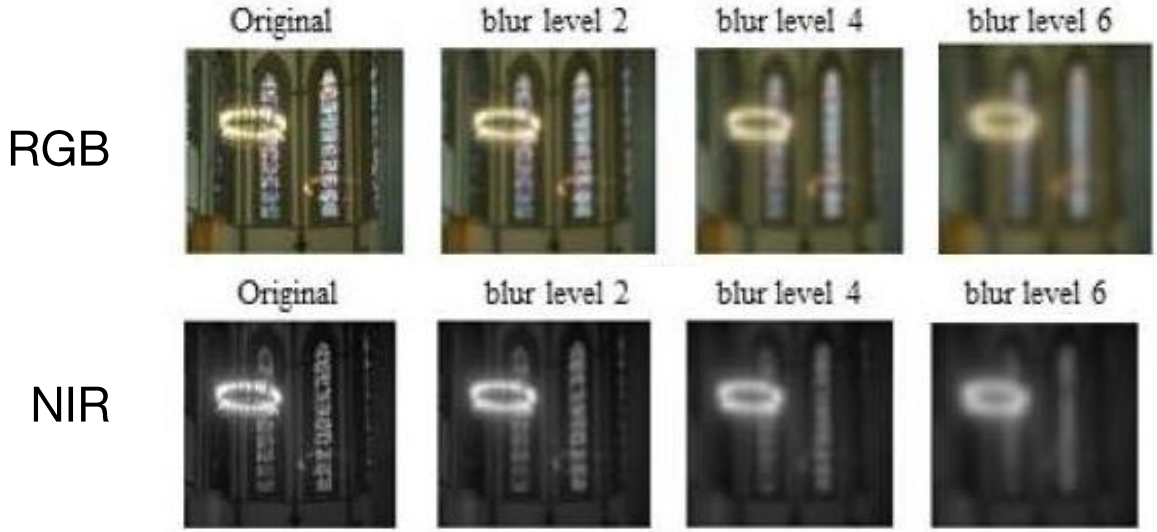
L. Karam et al., "Generative Sensing: Transforming Unreliable Sensor Data for Reliable Recognition," IEEE MIPR, 2018.

# Face Recognition using the SCface Dataset: DNN Feature Maps



# Scene Recognition using the EPFL RGB-NIR Scene Dataset

- The EPFL RGB-NIR Scene dataset consists of 9 scene categories with at least 50 images per class, for both visible (RGB) and near-infrared spectra (NIR).




M. Brown and S. Süsstrunk, Multispectral SIFT for Scene Category Recognition, IEEE International Conference on Computer Vision and Pattern Recognition (CVPR), pp. 177-184, 2011.

# Top-1 Accuracy Results for Scene Recognition using the EPFL RGB-NIR Scene Dataset.

| Method-Modality | Sensor Resolution: Blur Level | | | | | | | Avg |
|---|---|---|---|---|---|---|---|---|
| | $\sigma_b = 0$ | $\sigma_b = 1$ | $\sigma_b = 2$ | $\sigma_b = 3$ | $\sigma_b = 4$ | $\sigma_b = 5$ | $\sigma_b = 6$ | |
| Baseline-RGB | **0.9444** | 0.8466 | 0.7644 | 0.6177 | 0.4622 | 0.3511 | 0.2911 | 0.6110 |
| *Generative Sensing*- RGB | 0.9333 | **0.8555** | **0.8511** | **0.8555** | **0.8333** | **0.8355** | **0.8200** | **0.8548** |
| Baseline-NIR | **0.7629** | 0.6733 | 0.5911 | 0.4000 | 0.3088 | 0.2488 | 0.2200 | 0.4578 |
| *Generative Sensing*-NIR | 0.7518 | **0.7200** | **0.7177** | **0.6977** | **0.6622** | **0.6377** | **0.6222** | **0.6870** |

L. Karam et al., "Generative Sensing: Transforming Unreliable Sensor Data for Reliable Recognition," IEEE MIPR, 2018.

# Concluding Remarks

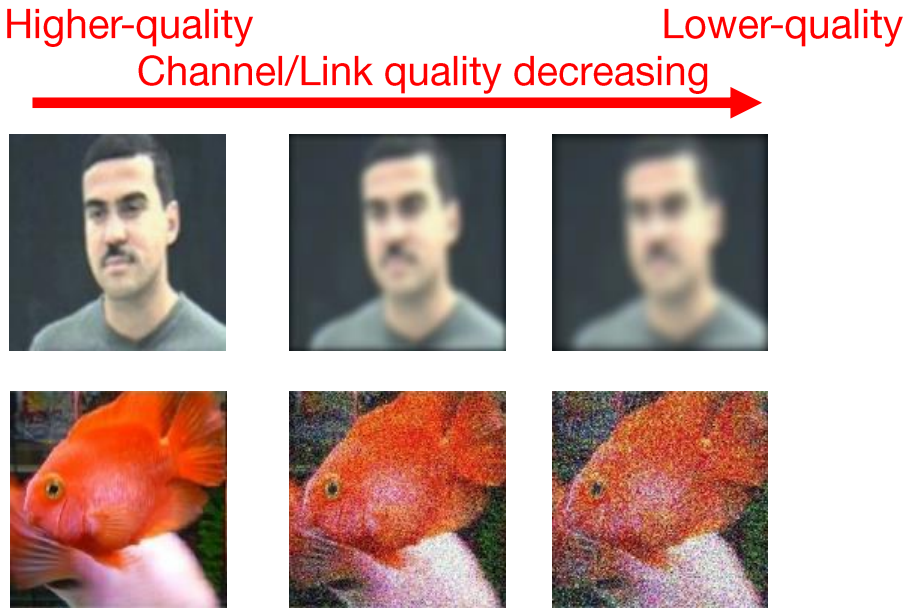
- Despite the generative units being trained using ImageNet (object recognition, RGB), they can be applied without retraining to other tasks (face recognition, scene recognition) and modalities (IR, NIR) with significant improvement in classification performance.
- For the SCface dataset, generative sensing features outperformed the baseline features with a 103% and 105% relative improvement in mean accuracy for the visible and infrared spectrum, respectively.
- For the RGB-NIR Scene dataset, generative sensing features outperform the baseline features with a 40% and 50% relative improvement in mean accuracy for the visible and near-infrared spectrum, respectively.
- The large performance gap between generative sensing features and the baseline features highlights the generic nature of our modality-invariant and sensor resolution-invariant features learnt by generative sensing models.

# Future Directions

- Construction of large datasets with matched data from various types of sensors (RGB, IR, Radar, Ultrasound,…)
- Demonstration of generative sensing with actual training of generative units with various sensing modalities.
- Investigation of other learning models for generative units.
- Development of generative sensing based robust low-power and low-cost sensing platforms that can work under varying conditions without compromising the recognition performance.
- Extension to communications/transmission.

# From Acquisition to Transmission

- Similar concepts can be translated from acquisition (sensors) to transmission
- Data quality impacted by quality of transmission medium (channel/network)



# Thank You

IVU Lab PhD Students

Tejas Borkar

Samuel Dodge

# Thank You

http://lina.faculty.asu.edu/

karam@asu.edu