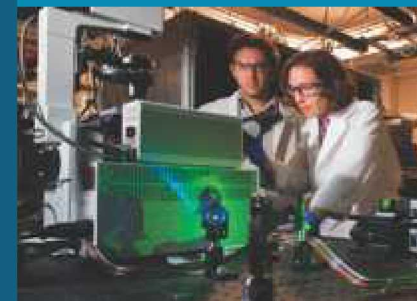
SAND2018-12580PE

Falls Risk Classification Using Smartphone Based Inertial Sensors and Deep Learning



PRESENTED BY

Matthew Martinez

Sandia National Laboratories is a multimission laboratory managed and operated by National Technology & Engineering Solutions of Sandia, LLC, a wholly owned subsidiary of Honeywell International Inc., for the U.S. Department of Energy's National Nuclear Security Administration under contract DE-NA0003525.

INTRODUCTION

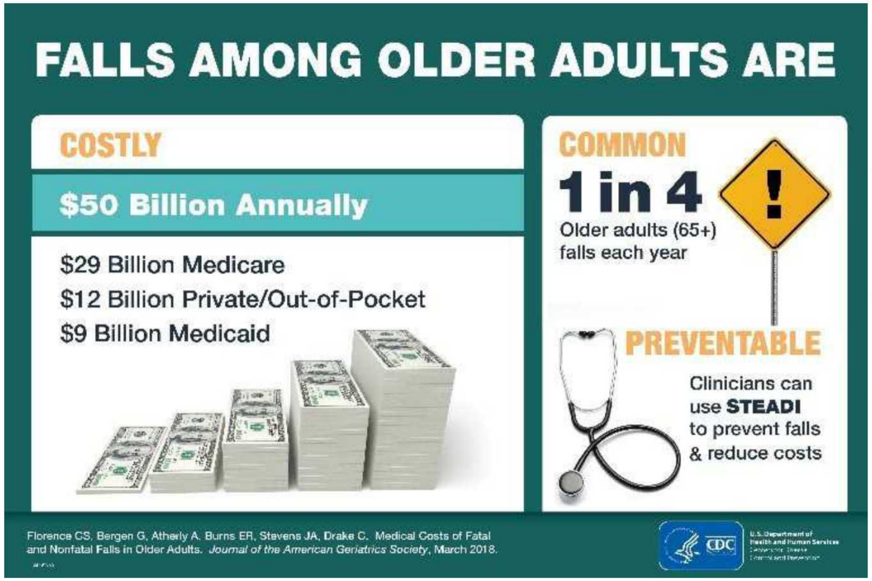
Falling in Older Adults

Definition of a fall
- A fall is an event that causes a person to rest inadvertently on the floor or other lower level

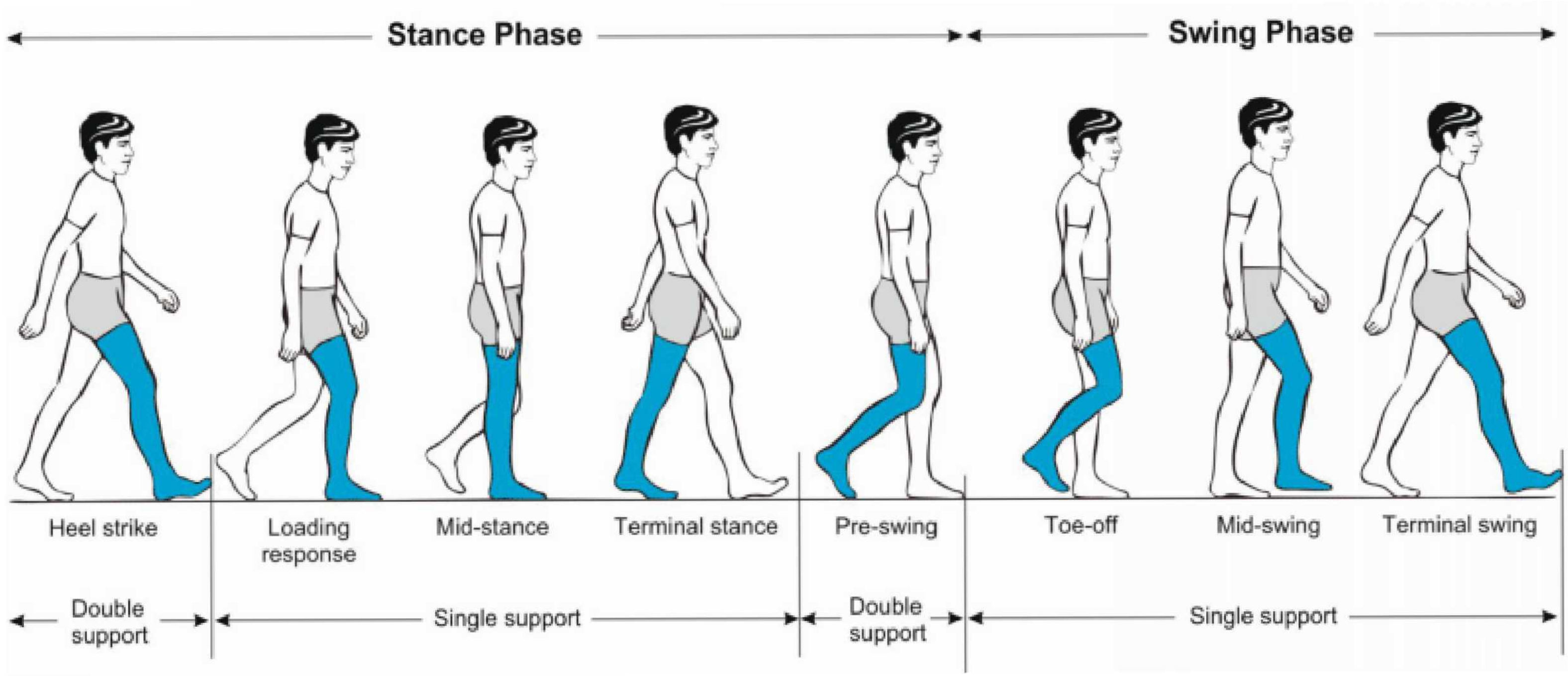
Each year 2.8 million adults are treated for fall related injuries
- Broken bones, hip fractures, traumatic brain injury
- Results in 800,000 hospitalization each year
- Medical costs exceed $50 billion

Emotional Cost of falling
- Increase fear of falling
- Decline in physical activity
- Reduced social interactions
- Depression

Falls Prevention Research
- Research has focused on assessment, prevention, and rehabilitation
- Prior research has focused on factors that attribute to falling
- Qualitative- and mobility-based assessments
- Sensor systems for monitoring gait
- Machine learning for gait analysis

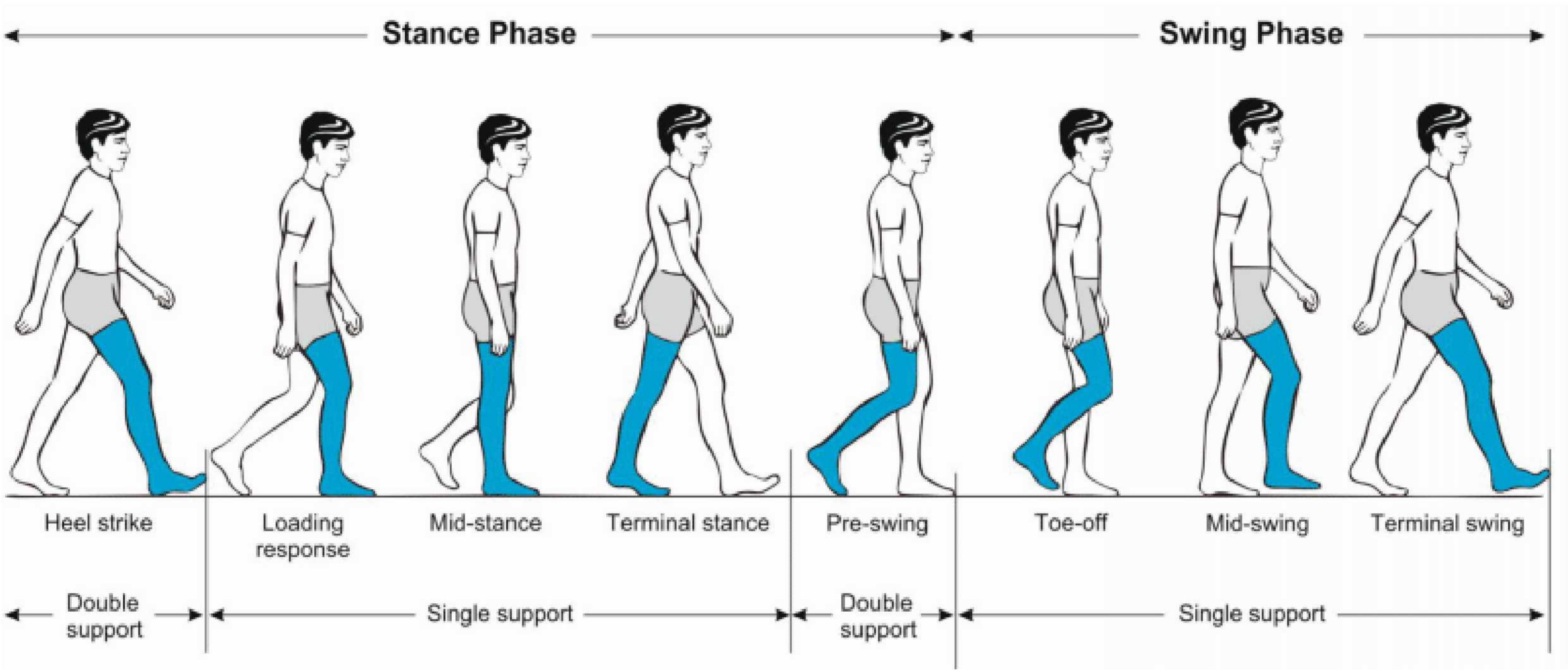


Gait Cycle



Stance Phase
Swing Phase
Heel strike
Loading response
Mid-stance
Terminal stance
Pre-swing
Toe-off
Mid-swing
Terminal swing
Double support
Single support
Double support
Single support

Gait Cycle



The 5 gait variables associated with falls risk

- Gait speed (cm/sec), cadence (step/min), stride length(cm)
- Swing time (%GC), double support time (%GC)

The 5 gait variables will be used to determine level of risk later

The Risk of Falling

Extrinsic Risk Factors

- Medications: psychotropics, diabetes medication, cardiovascular medication
- Home environment: Poor lighting, loose rugs, etc.
- Footwear: use of slippers and walking barefoot

Intrinsic Risk Factors

- Demographics, Bodily system functioning, Disease associated symptoms

Demographics

- Adults over 85 fall 4x more than those between 65 and 74
- Women are 58% more likely than men to suffer a non-fatal fall
- Men are 46% more likely to experience a fatal fall
- White women are 2.5x more likely than African American women to experience a fatal fall
- White women have higher incidence of fall-related hip fractures than African American women

System Decline

- Gait and balance disorders, Decrease in strength, Decline in vision

Disease Associated Symptoms

- Dizziness, Vertigo, Cardiovascular disease, Dementia, Depression

Gait and Balance disorders are one of the strongest indicators of risk

Sensors for Gait Analysis

Sensor for Gait Analysis
- 3-D motion capture
- Pressure sensitive walkways
- Inertial sensors

Smartphones for gait measurement
- Suite of sensors ideal for monitoring falls risk
- Microelectromechanical Systems (MEMS) inertial measurement units
  - 6- or 9-axis inertial sensors
- Open development environment
- Powerful processing capabilities
  - Mobile Machine Learning and Deep Learning APIs
- Continuous gait monitoring
  - In-home gait monitoring
  - Removes the need for domain experts to analysis test results

MEMS sensors provide a low cost solution for gait monitoring

Identifying at Risk Older Adults

Methods for Label Assignment:
- Survey based methods
- Clinical based methods

Survey Based Methods:
- Falls risk screenings (CDC STEADI)
- Comprehensive Falls Risk Screening Instrument (CFRSI) [Multi-component method]

Clinical Based Methods:
- Timed Up & Go:
  - Individuals who take > 12 s are considered at risk of falling
- 30-sec Chair Stand:
  - Individuals whose score is below a given threshold are at risk of falling,
  - age > 75 threshold is 11 (Men) and 10 (Women)
- 4-Stage Balance Test:
  - Individual's unable to hold a tandem stand for 10s is at risk of falling

https://www.cdc.gov/steadi/materials.html

Features for Falls Prediction

Features

- **Position and Angle**
  - Peak-to-Peak Amplitude
  - Trunk Tilt
  - Postural Sway During Stance
- **Angular Velocity**
  - Max
  - Mean
  - Peak-to-Peak Amplitude
  - Min
  - RMS During Stance
  - 3-D RMS During Stance
  - Signal Variability
- **Linear Acceleration**
  - Median
  - Std. Dev.
  - Peak-to-Peak Amplitude
  - RMS
  - 3-D RMS
  - RMS During Stance
  - 3-D RMS During Stance
- **Spatial Gait Parameters**
  - No. Steps
  - Step Length
- **Energy**
  - Local Wavelet Energy
- **Temporal Gait Parameters**
  - Gait Speed
  - Std. Dev. Stride Time
  - Step Duration
  - Step Duration Variability
  - Cadence
  - Stride Time
  - % Time Spent in Double Support
- **Spectral**
  - Spectral Edge Frequency
  - Ratio of Area Under 1st Harmonic to Remaining Area
  - Power Spectrum Energy
  - Ratio of first 4 Harmonics to first 6
  - No. FFT Peaks
  - 1st FFT Peak Parameters
  - 25%/50%/75% Quartile Frequency
  - Dominate FFT Peak
  - Ratio of Even & Odd Harmonics
  - Dominate FFT Peak Parameters

Spectral features are the most discriminative

Classifiers for Falls Prediction

Falls Classification Problems:

Falls Prediction → Classification using retrospective falls history for label assignment

- Faller/Non-Faller Labels

Prospective Falls Prediction → Classification that an individual will experience a future fall

- Prospective Faller/Non-Faller

Falls Risk Classification → Classification of risk of experiencing a future fall

- Low/High Risk Labels

Classifiers

- Logistic Regression (89.1% ACC), neural networks (MLP) (91% ACC), naïve Bayes (95% ACC), SVM (85% ACC)
- Prospective fallers: Random Forests (73.4% ACC)
- Best classifiers used frequency domain features

Feature Selection

- Correlation feature selection
- Fast correlation feature selection
- Relief-F
- Features selection provided 9% increase in model performance

Neural networks (MLP) and Naïve Bayes are the best classifiers

Anomaly Detection for Falls Risk Prediction

M. Martinez, P. L. De Leon, and D. Keeley, "Novelty Detection for Predicting Falls Risk using Smartphone Gait Data", in *Pro. IEEE Int. Conf. Acoustics, Speech & Signal Proc.* (ICASSP), 2017

Dataset
- Walkway measurements of biomechanical data
- Inertial measurement based

Data Labeling
- Used walkway data to determine low and high risk labels
- Low/high risk labels were based on independent thresholds risk ratios

Features
- Extracted 21 features from the harmonic spectrum
- Fundamental frequency
- Ratio of the area under the 1st harmonic to the area under the first 6 harmonics
- Ratio of the area under the first 6 harmonics to the total spectrum
- Ratio of the area under the even harmonics to the area under the odd harmonics

Analyzed Anomaly Detection Methods
- One-Class Support Vector Machine (Hyperplane)
- Support Vector Data Decision (Sphere)

Results: Achieved F1 score of 79.07% for SVDD with RBF kernel

Improvement needed for data labeling, feature extraction, and classification

Motivations and Contributions

Challenges in Falls Prediction Research

- Feature engineering
- Retrospective falls history
- Models trained on low number of examples

Contributions

- Propose using gait variables associated with an increase risk of falling to provide label assignment
- Propose using deep neural networks for learning features related to human motion
- Apply transfer learning to adapt a pre-trained network for falls risk classification

Summary of Dissertation Results

- We demonstrate by using *k*-means we can cluster vectors of risk ratio as low/high falls risk
- We show that by using a Bayesian classifier we can classify vectors of gait variables as low/high risk while quantifying classifier uncertainty
- We show how to pre-train a deep neural network to learn feature representation related to human motion using publicly available pedestrian activity data
- We show how to use a pre-trained deep neural network as feature extractor for falls risk classification
- We show how to classify falls risk using inertial gait measurements collected from a smartphone

Older Adult Walkway Data

Older Adult Walkway Data

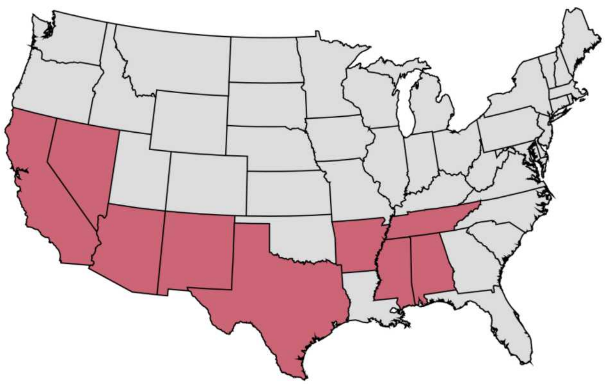
Data collected in partnership with the Electronic Caregiver Company

Sensor System:

- TekScan® Walkway™ System
- Measures plantar force and pressure
- Automatic foot strike detection, foot labeling, segmentation
- Gait Variables, i.e. gait speed, cadence, stride length, swing time, and 2x support

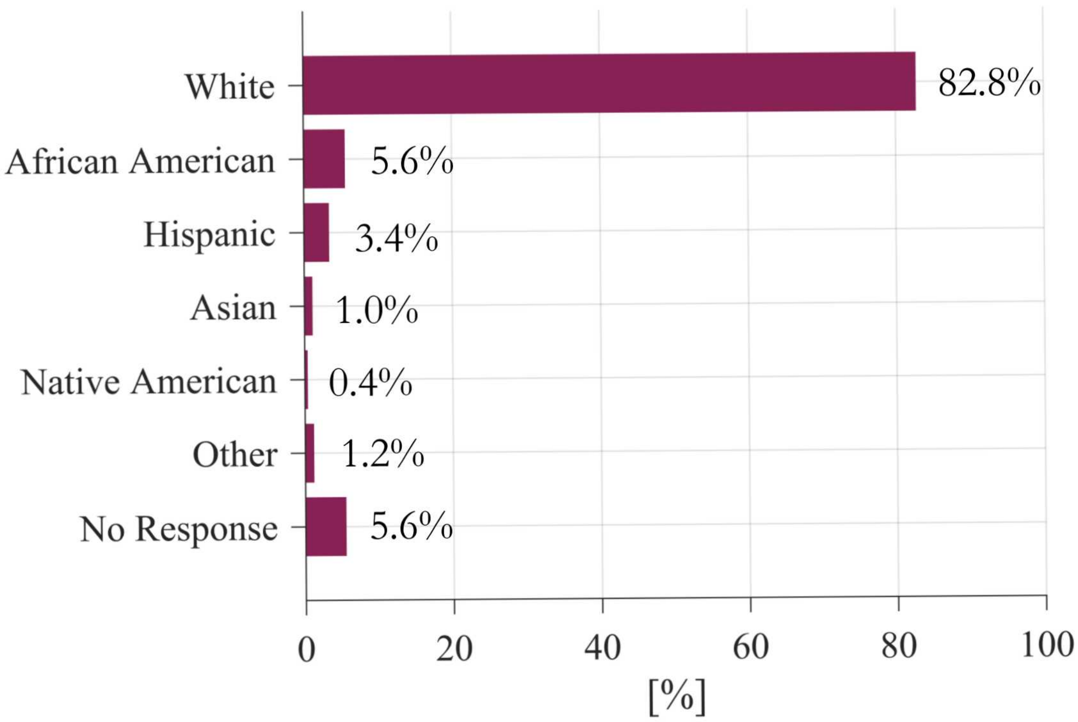
Data Collection:

- Data collected from 854 participants
- Collected at 50 in Southern and Southwestern US
- Participant Selection Criteria
  - Age, cognitive ability
  - Ability to read and understand liability waiver
  - Ability to ambulate for 30s w/ or w/o assistive device



Use of gait datasets approved for secondary analysis by NMSU IRB reference #15405

Older Adult Walkway Data



All ethnicities except Whites are under represented

Older Adult Walkway Data

Variable	Female (n=601)	Male (n=253)
Age, y (std. dev.)	78.9 (7.9)	78.5 (7.8)
Weight, kg (std. dev.)	70.8 (16.5)	85.3 (17.4)
Cardiovascular (%)	237 (38.4)	91 (36.0)
Arthritis (%)	323 (53.7)	109 (43.1)
Neurologic (%)	183 (30.4)	68 (26.9)
Metabolic (%)	213 (35.4)	83 (32.8)
More-than-one (%)	248 (41.3)	89 (35.2)

Female/Male proportion is than the U.S. population

Healthier than the overall U.S. over the age of 65

BAYESIAN CLASSIFICATION FOR FALLS RISK USING WALKWAY DATA

M. Martinez, P. L. De Leon, and D. Keeley, “Bayesian Classification for Falls Risk”, *Gait & Posture*, vol. 67, pp. 99 - 103, Jan 2019

Falls Risk Labeling

Falls Risk Labels

- Labels obtained from questionnaires → falls history, medication usage, home environment
- Clinical methods → Timed Up & Go, 30-s chair stand, 4-stage balance test

Challenges with questionnaires and clinical tests

- Questionnaires are error prone
- Thresholds to determine faller and non-faller
- Faller/Non-faller labels are indicative of an individual's falls history
- Do not account for biomechanical factors associated with a prospective fall
- Clinical test do not quantify the uncertainty in the classifier's decision
- Clinical methods do not provide a probability of adverse outcome

Bayesian Classifier

- Develop a Bayesian classifier for classifying older adults as low or high risk
- Walkway data will be used to train classifier
- Calculate risk vectors that quantify how falls risk changes with gait variables
- Risk vectors are used to train *k*-means
- Gait vectors are used to train a two-component GMM
- GMM calculated posterior probabilities are used to build Bayesian classifier
- Monte Carlo simulation to assess classifier performance

Biomechanical Risk Factors

Age related gait degradation

- Stiffer and less coordinated gait pattern
- Less capable of self-correcting after a slip or trip
- Decrease in muscle strength tone,
- Decreased step height and length
- Reduction in body orienting reflexes
- Unable to rapidly and correct balance

Gait Variables Related To Risk

- 5 Variables
- Pace factor (3 variables)
- Rhythm factor (2 variables)

Risk Ratios:

- Ratio of probability of exposed group to unexposed group
- RR = 1 → outcome is not effected by exposure
- RR < 1 → outcome is decreased by exposure
- RR > 1 → outcome is increased by exposure

Biomechanical Risk Factors

Age related gait degradation

- Stiffer and less coordinated gait pattern
- Less capable of self-correcting after a slip or trip
- Decrease in muscle strength tone,
- Decreased step height and length
- Reduction in body orienting reflexes
- Unable to rapidly and correct balance

Gait Variables Related To Risk

- 5 Variables
- Pace factor (3 variables)
- Rhythm factor (2 variables)

Risk Ratios:

- Ratio of probability of exposed group to unexposed group
- RR = 1 → outcome is not effected by exposure
- RR < 1 → outcome is decreased by exposure
- RR > 1 → outcome is increased by exposure

Gait Variable	Median	Unit Change	Risk Ratio
Gait Speed	95.1 cm/sec	-10 cm	1.078
Cadence	101.8 steps/min	-10 steps/min	1.085
Stride Length	112.5 cm	-10 cm	1.095
Swing Phase	36.6 %	-10 %	1.503
Double Support	26.6%	+10 %	1.207

Data Labeling Using Risk Ratios

Challenge with Risk Ratios

- Quantify how risk changes with changes in gait variables
- No method for calculating falls risk
- Completely unlabeld data set

Apply *k*-means to derive risk based labels

- First, calculate change in risk

$$\Delta_{risk} = \frac{x - \text{median}}{\text{unit change}} \times (\text{risk ratio} - 1)$$

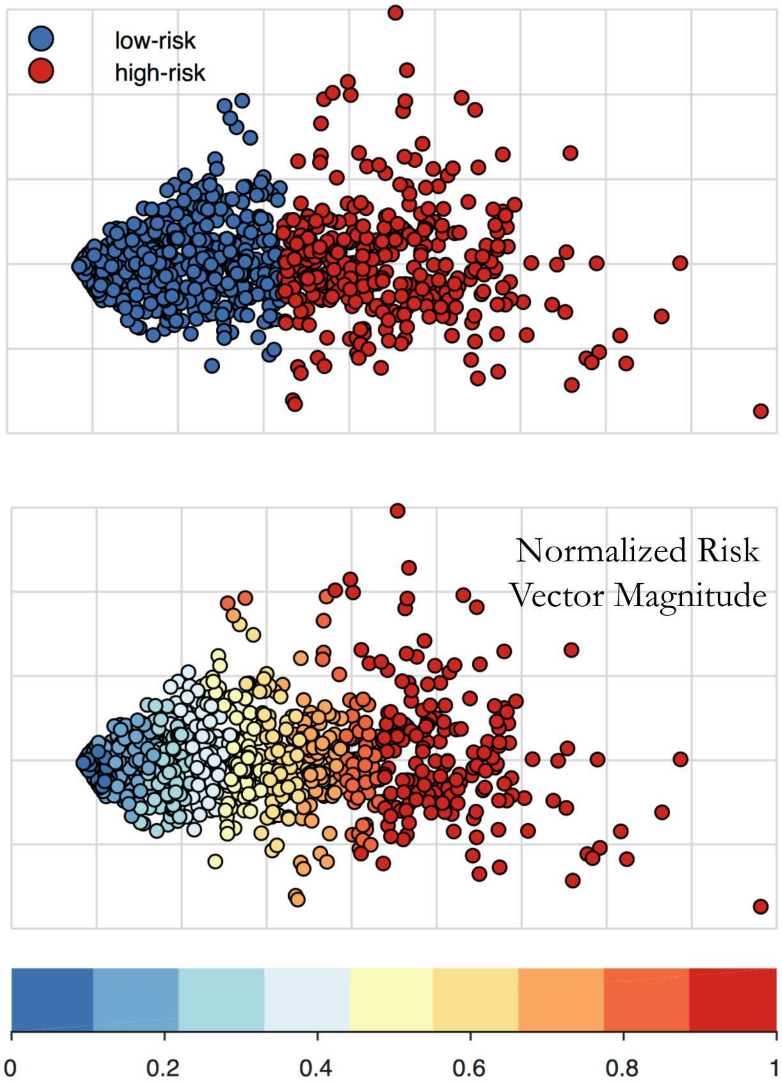
- $\Delta_{risk} < 0$, then $\Delta_{risk} = 0$
- Cluster into two classes (low/high risk)
  - assume that data structure consists of two classes
- Hard class decision or label assignment

k-means labels

- Use to risk labels to evaluate Bayesian classifier

Results

- Low Risk: 511 (59.8%) High Risk: 343 (40.2%)



Risk vectors are effectively clustered into low and high risk groups

Bayesian Classification

Bayesian Classifier

- Simple classifier constructed by Bayes Rule: $p(C|x) = \frac{p(x|C)p(C)}{p(x)}$

Training a Bayesian Classifier

- Learn a probability model for the likelihood
- Model class prior as equally likely or estimate class prior from data

Evaluating Classifier

- Maximum A Posteriori (MAP) decision
- Threshold using receiver operating characteristic analysis

Semi-supervised and Unsupervised Bayesian Classifier

- Assume that data is generate from a mixture distribution
- Number of components is equal to the number of classes
- Learning class structure through clustering algorithm

Class structure is learned through mixture models

Gaussian Mixture Modeling

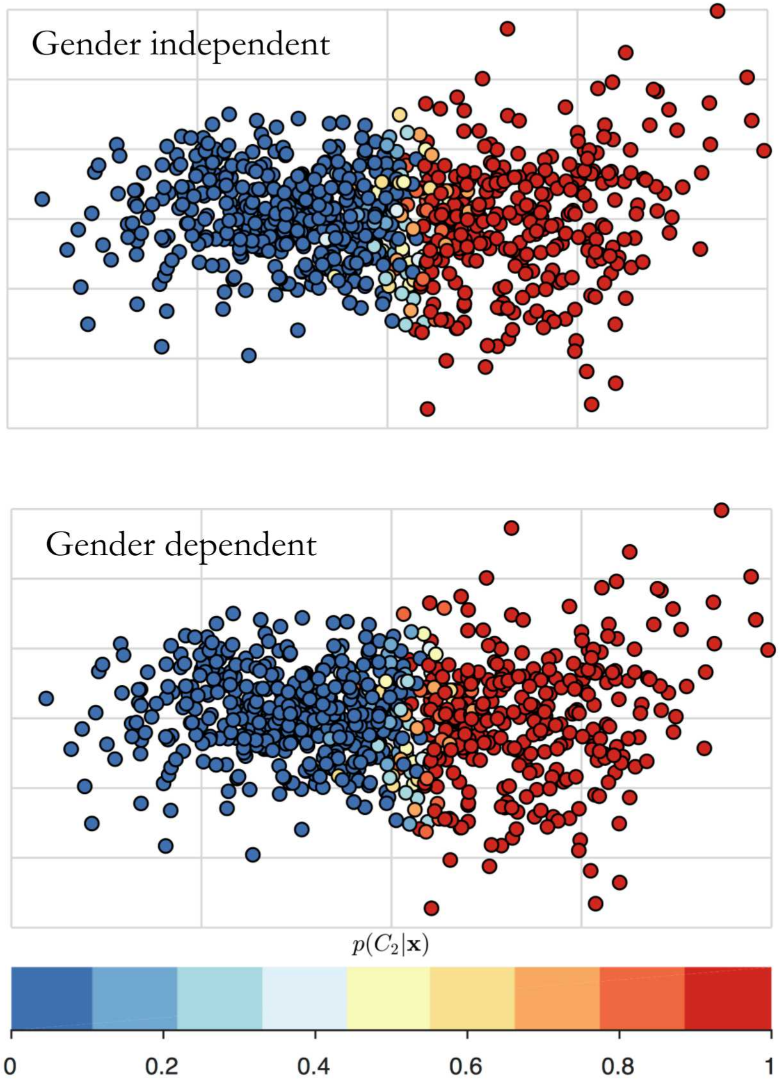
Mixture Modeling
- Model data using 2-component GMM
- Components represent low/high risk

GMM Parameter Estimation
- Use Expectation-Maximization algorithm
- Training Data: Vector of 5 gait variables
- Gait variables measured from walkway measurements

Model Training:
- Diagonal covariance matrix
- Low risk component is determined using median values for gait vector

Train 2 different models
- Gender independent model
- Gender dependent model

Soft label assignment



Risk assessment is performed using underlying biomechanics

Bayesian Classifier and Assessment Methodology

Bayesian Classifier

$$\begin{cases} C_1, & p(C_1|\mathbf{x}) \geq \theta \\ C_2, & \text{otherwise} \end{cases}$$

Assessment

- Every walkway example has two labels: 1). Risk based label and 2). GMM soft label
- Boot-strap the risk based labels (*k*-means) to evaluate Bayesian classifier
- Apply threshold to posterior probabilities

Assessment

- Optimal threshold is computed by maximizing Youden's J Statistic

$$J = \text{sensitivity} + \text{specificity} - 1$$

  - Gives equal weight to false positive and false negatives
- Sensitivity and specificity obtained from ROC curve

Monte Carlo Simulation (100 trials)

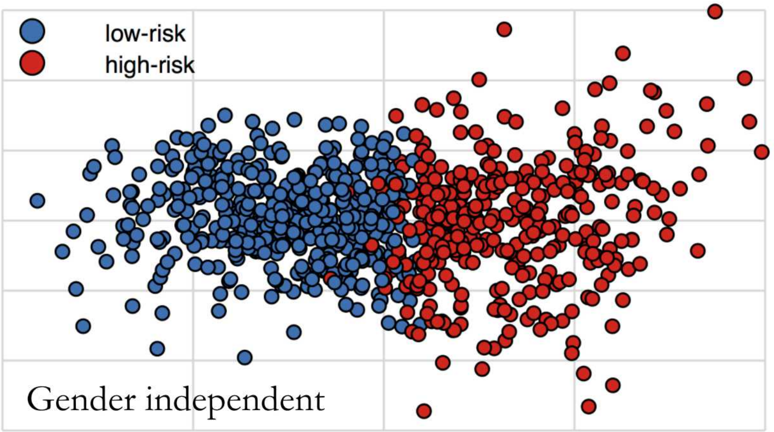
- 80% used to train, 20% validation
- Stratified re-sampling scheme

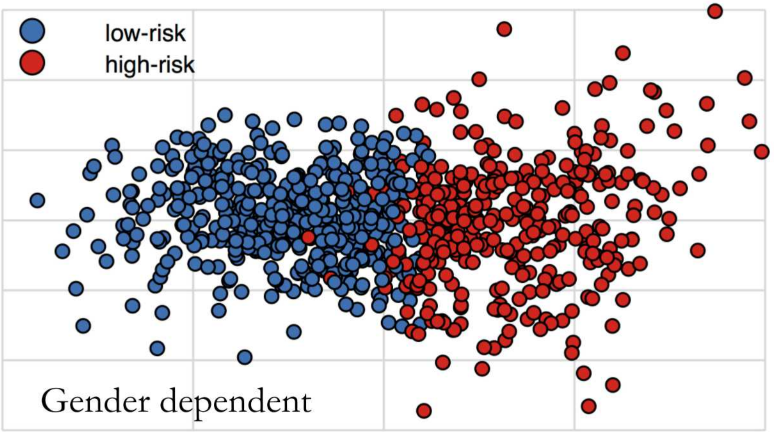
Model	**AUC-ROC**	$\boldsymbol{\theta}$	**ACC**	**SPEC**	**SENS**
Gender-indep	99.1%	0.461	96.5%	95.4%	98.1%
Female	99.1%	0.464	96.8%	95.3%	98.7%
Male	99.1%	0.601	95.5%	95.5%	95.4%

Results

Final Model:

- Retrain on all data
- Use thresholds from MC simulation
- ACC:
  - Gender independent 96.4%
  - Gender dependent: 97% (female) 96.8% (male)
- Minor improvement from gender modeling
- High agreement between hard labels and GMM labels
- Gender Dependent: 495 low, 359 high
- Gender Independent: 492 low, 362 high





Limitations:

- Walkway data collected in laboratory setting
- Self-reported health data
  - Anonymity, biased statement avoidance, request for truthful answers

Gait vectors can effectively be classified according to risk level

Summary

Demonstrated method for classifying older adults as low or high risk

- Bayesian classifier
- Posteriors probabilities were obtained from a two-component GMM
- Using walkway data GMM parameters were estimated for gender independent and dependent models.

Demonstrated good agreement with risk ratio based labels

- ROC curve analysis
- Accuracies greater than 96%

Clinical Use

- Can be used to assist clinicians with identifying older adults' falls risk using gait data
- Can improve clinicians' recommendations for intervention

The Bayesian classifier provides a method for assigning labels to inertial data

FALLS RISK CLASSIFICATION VIA TRANSFER LEARNING

M. Martinez, and P. L. De Leon "Falls Risk Classification of Older Adults using Deep Neural Networks and Transfer Learning ", in Review *IEEE J. Biomed. And Health Inform*. Submitted Sept. 2018

Deep Learning for Biomedical

Machine Learning and Deep Learning for Biomedical

- Detection of influenza epidemics using search engine data
- Skin cancer classification using deep convolutional neural networks
- Diabetic retinopathy detection using deep convolutional neural networks

Deep Learning for Gait Analysis and Gait Disorder Classification

- Gait parameter estimation (stride length, stride width, swing time, etc) foot mounted inertial sensor data (CNN)
- Gait pattern classification from tomography sensor data (CNN)
- Detection of freezing of gait in Parkinson's patients (CNN)

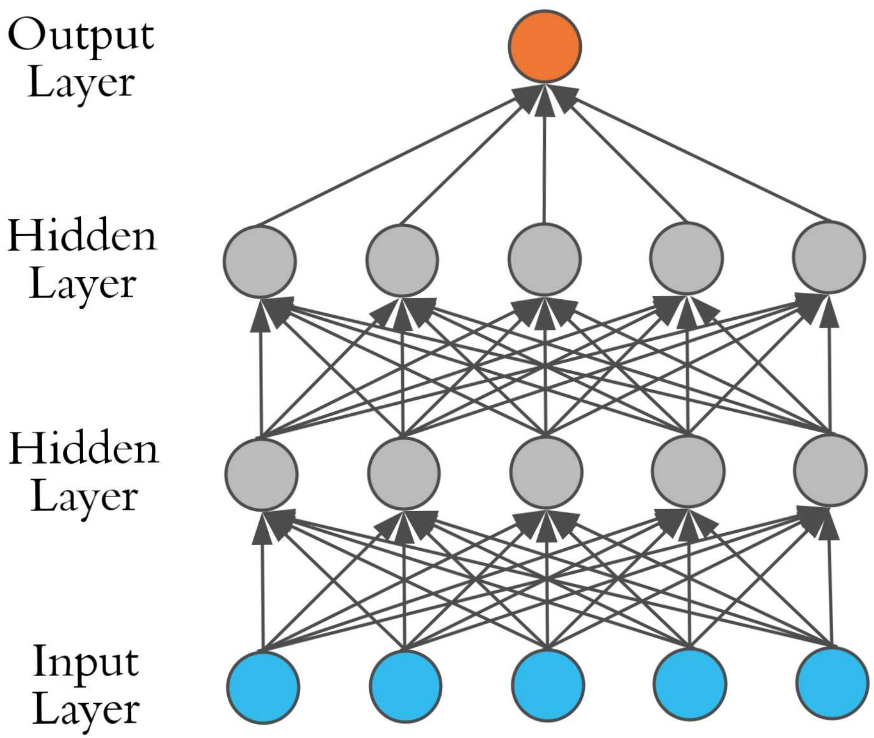
Deep Learning for Falls Risk Classification

- To our knowledge first to use deep neural networks for falls risk classification from inertial sensor data (6-axis)

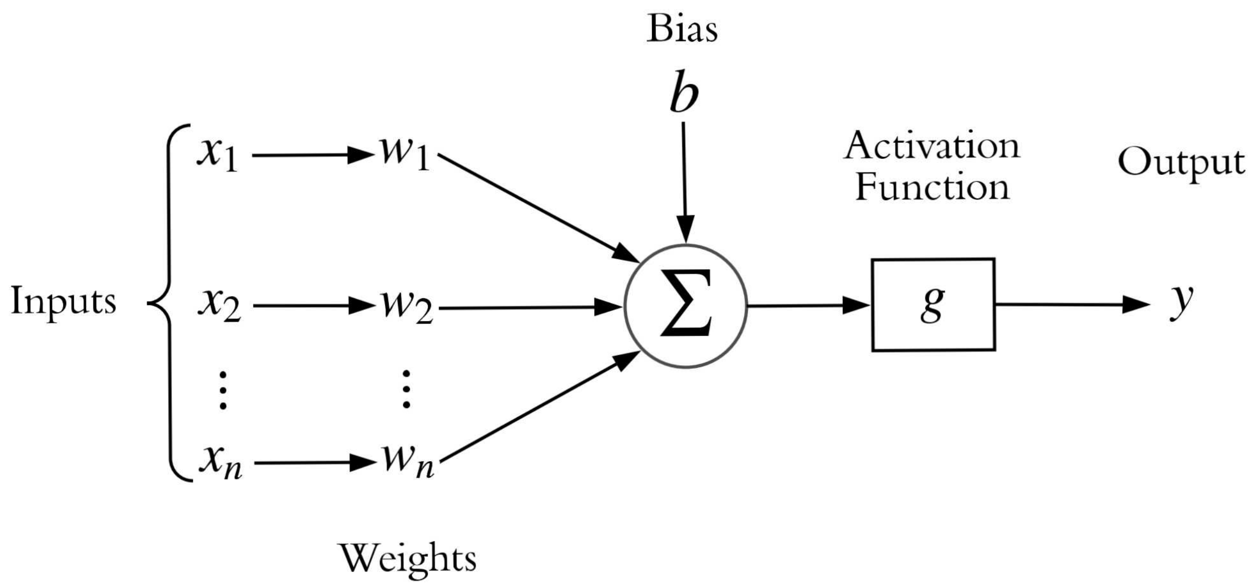
Proposed Method:

- Train a deep neural network to classify sequences of inertial gait data as low/high risk
- First train a fully convolutional neural network for a pedestrian activity recognition task
- Use transfer learning to solver the falls risk classification problem

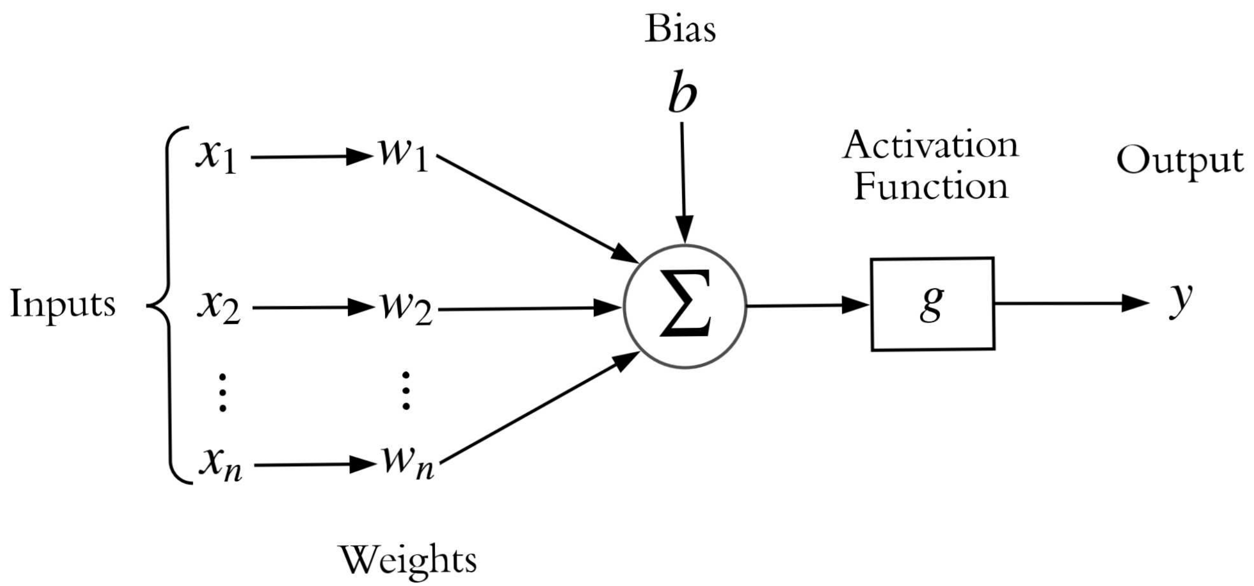
Neural Networks

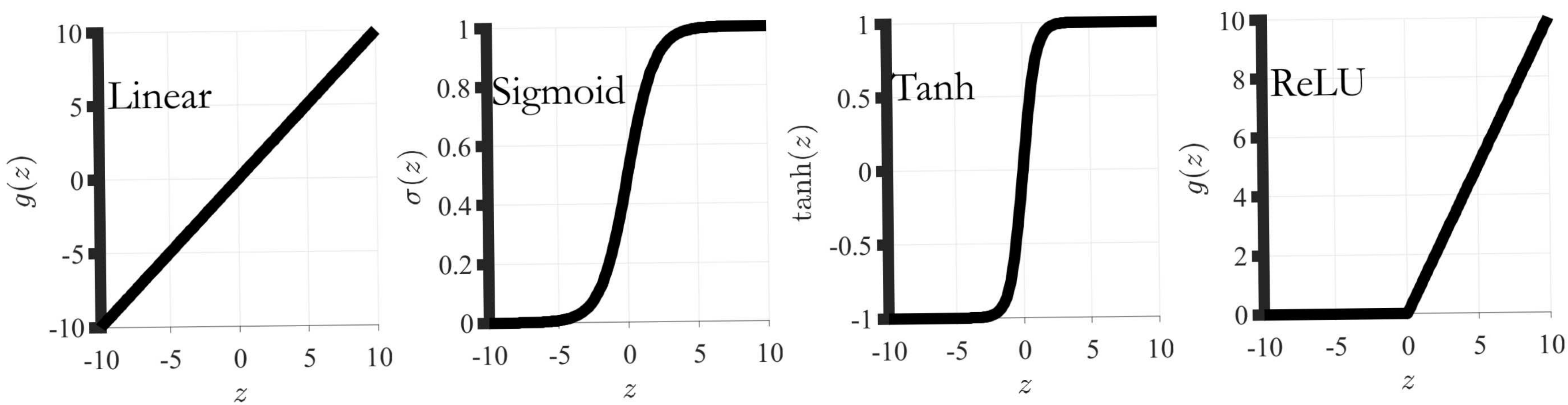


Neural Networks



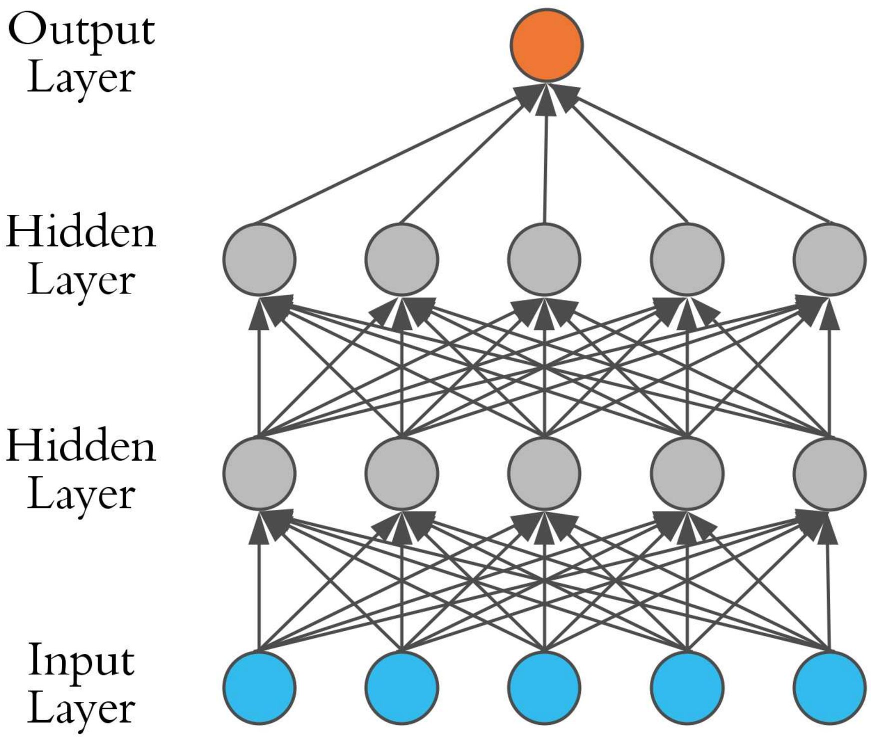
Neural Networks





Activation Functions

Neural Networks

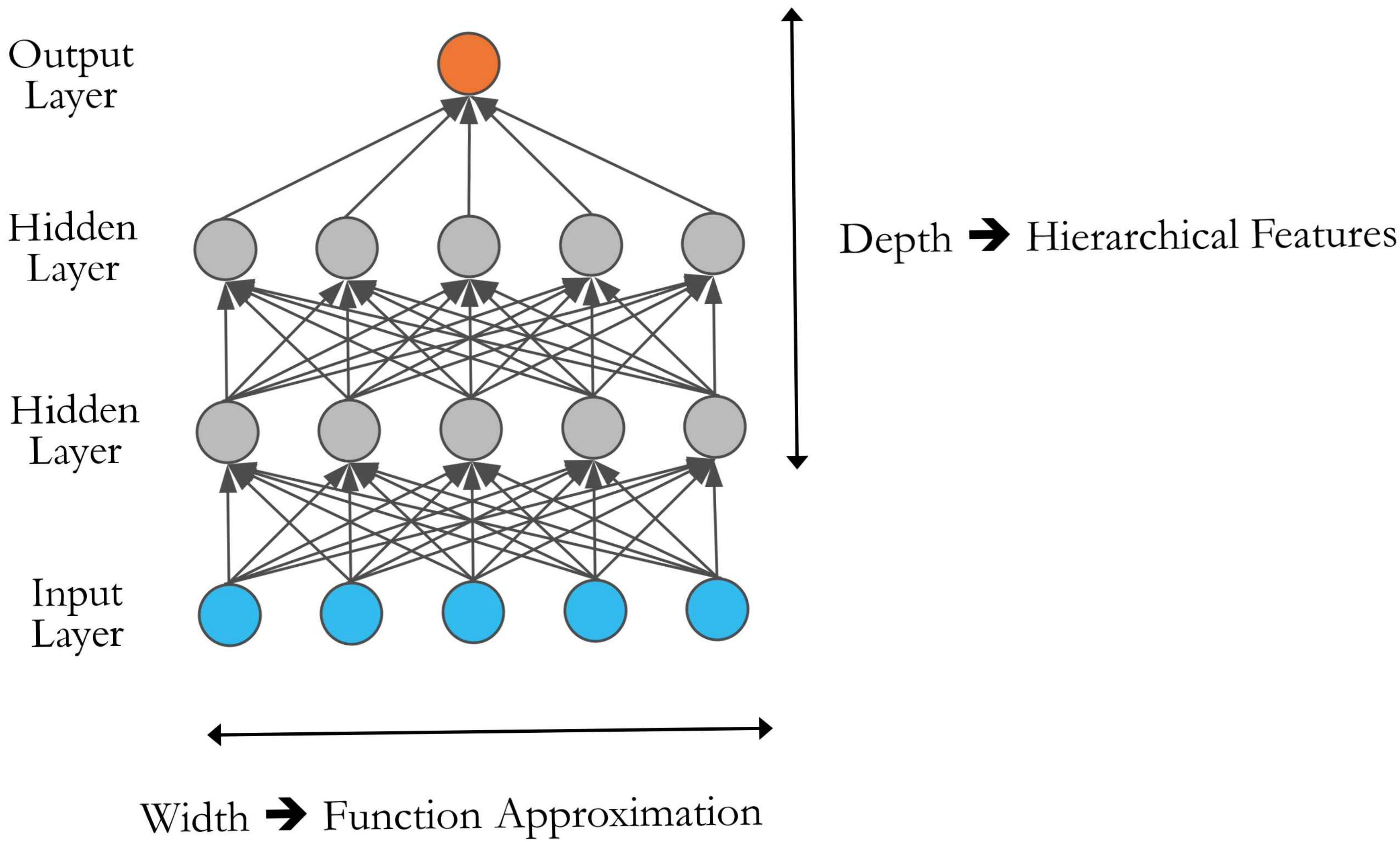


$$y = g\left(\mathbf{h}_2^T \mathbf{w}_o + b_o\right)$$

$$\mathbf{h}_2 = g\left(\mathbf{h}_1^T \mathbf{w}_2 + \mathbf{b}_2\right)$$

$$\mathbf{h}_1 = g\left(\mathbf{x}^T \mathbf{w}_1 + \mathbf{b}_1\right)$$

Neural Networks



Convolutional Neural Networks

CNN

- Specialized neural network for data with a grid like structure
  - 1-D (time series), 2-D (images), or 3-D (volumetric images)
- CNNs are more computationally efficient (less trainable parameters)
- Matrix multiplication is replaced with convolution

$$\mathbf{h} = g(\mathbf{x} * \mathbf{W} + \mathbf{b})$$

- Convolution allows network to process data of variable sizes

Convolutional Neural Networks

CNN

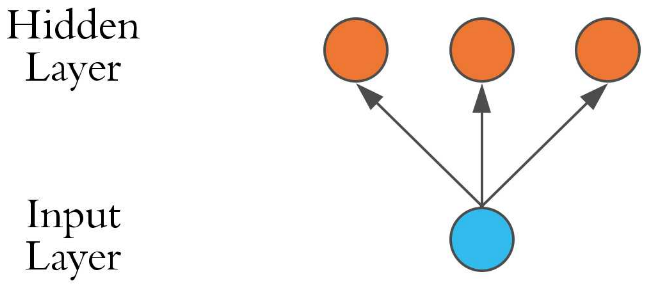
- Specialized neural network for data with a grid like structure
  - 1-D (time series), 2-D (images), or 3-D (volumetric images)
- CNNs are more computationally efficient (less trainable parameters)
- Matrix multiplication is replaced with convolution

$$\mathbf{h} = g(\mathbf{x} * \mathbf{W} + \mathbf{b})$$

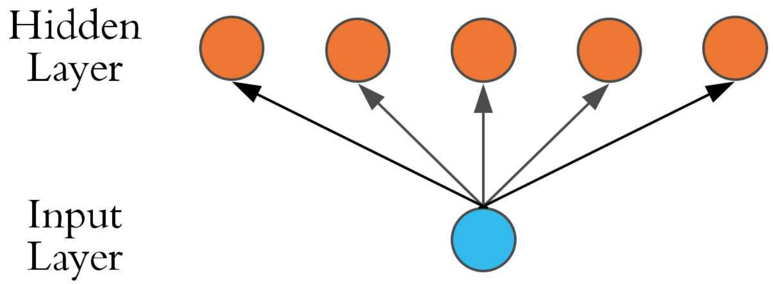
- Convolution allows network to process data of variable sizes

Properties of CNN

- 1). Sparse interactions



Adapted from "Deep Learning"



Adapted from "Deep Learning"

Convolutional Neural Networks

CNN

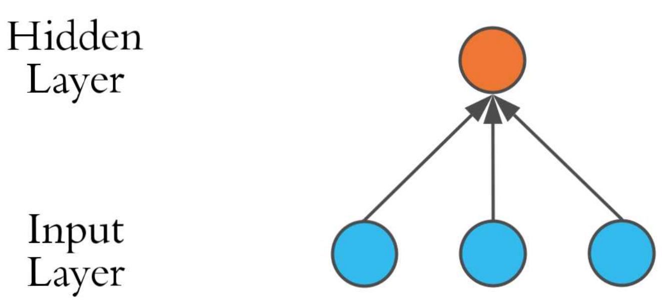
- Specialized neural network for data with a grid like structure
  - 1-D (time series), 2-D (images), or 3-D (volumetric images)
- CNNs are more computationally efficient (less trainable parameters)
- Matrix multiplication is replaced with convolution

$$\mathbf{h} = g(\mathbf{x} * \mathbf{W} + \mathbf{b})$$

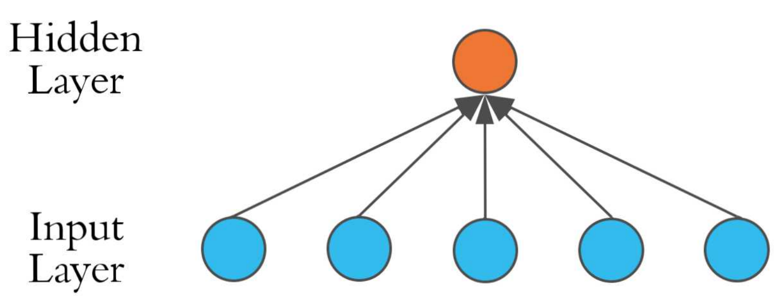
- Convolution allows network to process data of variable sizes

Properties of CNN

- 1). Sparse interactions → Computational and memory efficient



Adapted from "Deep Learning"



Adapted from "Deep Learning"

Convolutional Neural Networks

CNN

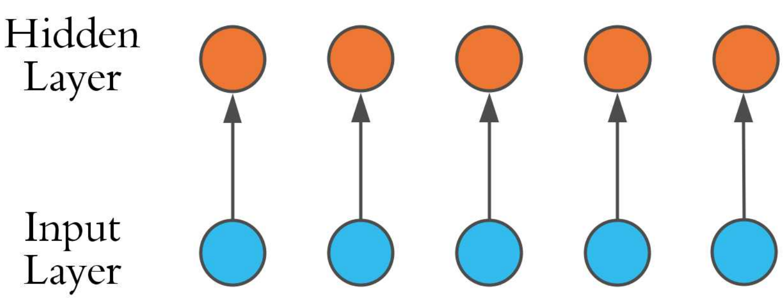
- Specialized neural network for data with a grid like structure
  - 1-D (time series), 2-D (images), or 3-D (volumetric images)
- CNNs are more computationally efficient (less trainable parameters)
- Matrix multiplication is replaced with convolution

$$\mathbf{h} = g(\mathbf{x} * \mathbf{W} + \mathbf{b})$$

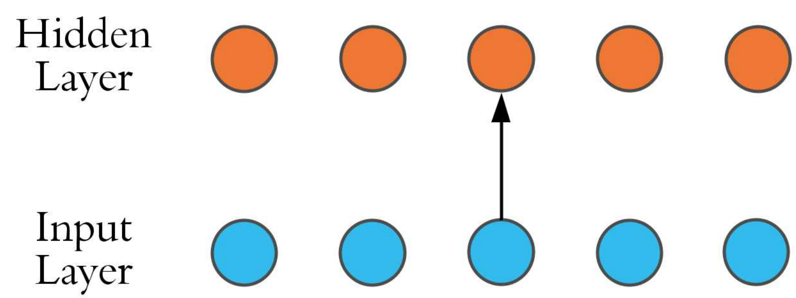
- Convolution allows network to process data of variable sizes

Properties of CNN

- 1). Sparse interactions → Computational and memory efficient
- 2). Parameter sharing → Improved memory efficiency



Adapted from "Deep Learning"



Adapted from "Deep Learning"

Convolutional Neural Networks

CNN

- Specialized neural network for data with a grid like structure
  - 1-D (time series), 2-D (images), or 3-D (volumetric images)
- CNNs are more computationally efficient (less trainable parameters)
- Matrix multiplication is replaced with convolution

$$\mathbf{h} = g(\mathbf{x} * \mathbf{W} + \mathbf{b})$$

- Convolution allows network to process data of variable sizes

Properties of CNN

- 1). Sparse interactions → Computational and memory efficient
- 2). Parameter sharing → Improved memory efficiency
- 3). Equivariant representations → When the input changes the output changes accordingly

Time series classification

Time Series Classification

- Classify a vector sequence, $\mathbf{x}_t$, to one of C classes
- Smartphone IMU data are treated as vector sequence
- Each element is a sample from an independent sensor channel

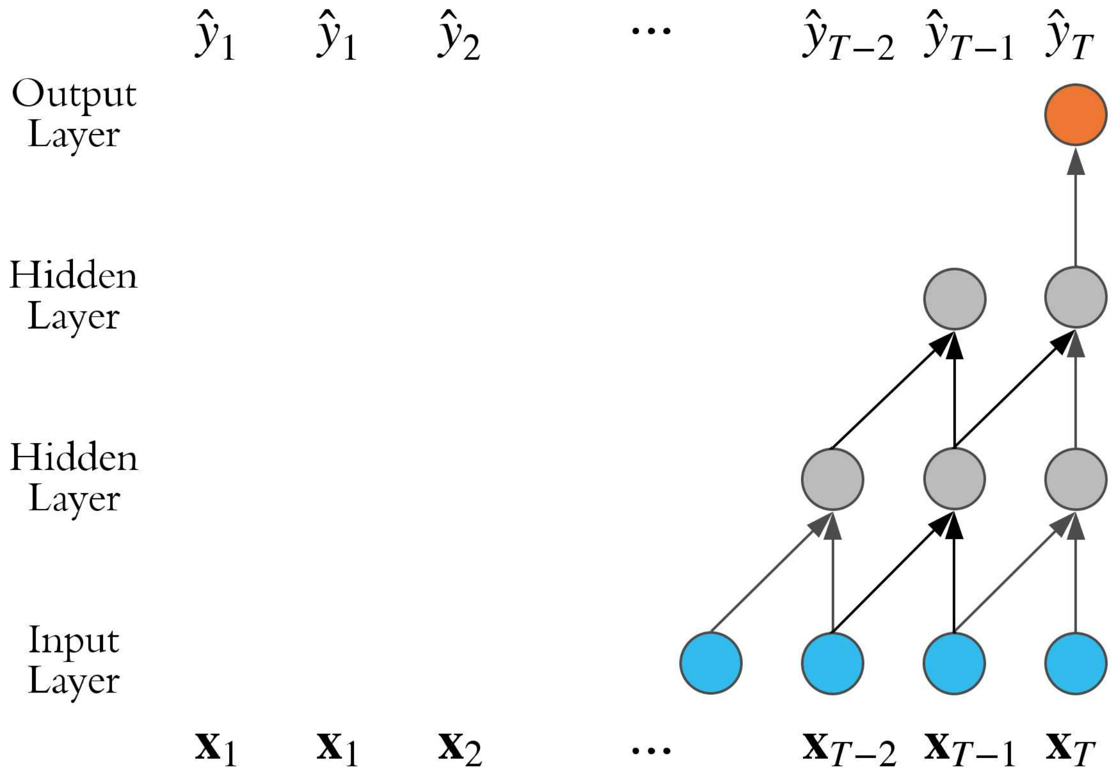
Human Activity Recognition

- Deep Convolutional Neural Networks
- Recurrent Neural Networks
- Sequence-to-sequence learning
- Hybrid CNN-RNN architectures

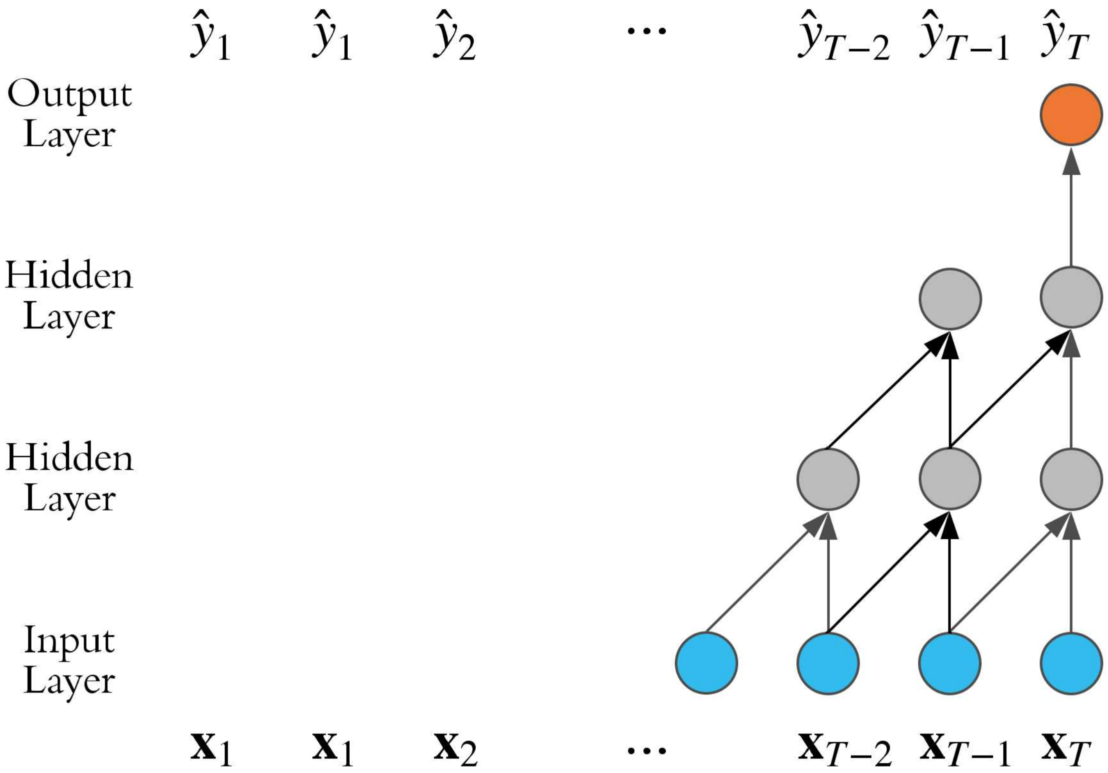
Fully Convolutional Neural Networks

- Typically used for semantic image segmentation
- Process arbitrary sequence since fully-connected layer is omitted
- Extract features at different time scales by using causal dilated convolutions
- FCNNs have outperformed RNNs on the same tasks
- Easier to train than RNNs
- Have less trainable parameters than RNNs
- Do not suffer from the vanishing/exploding gradient problem
- Achieved state-of-the-art results for time series classification

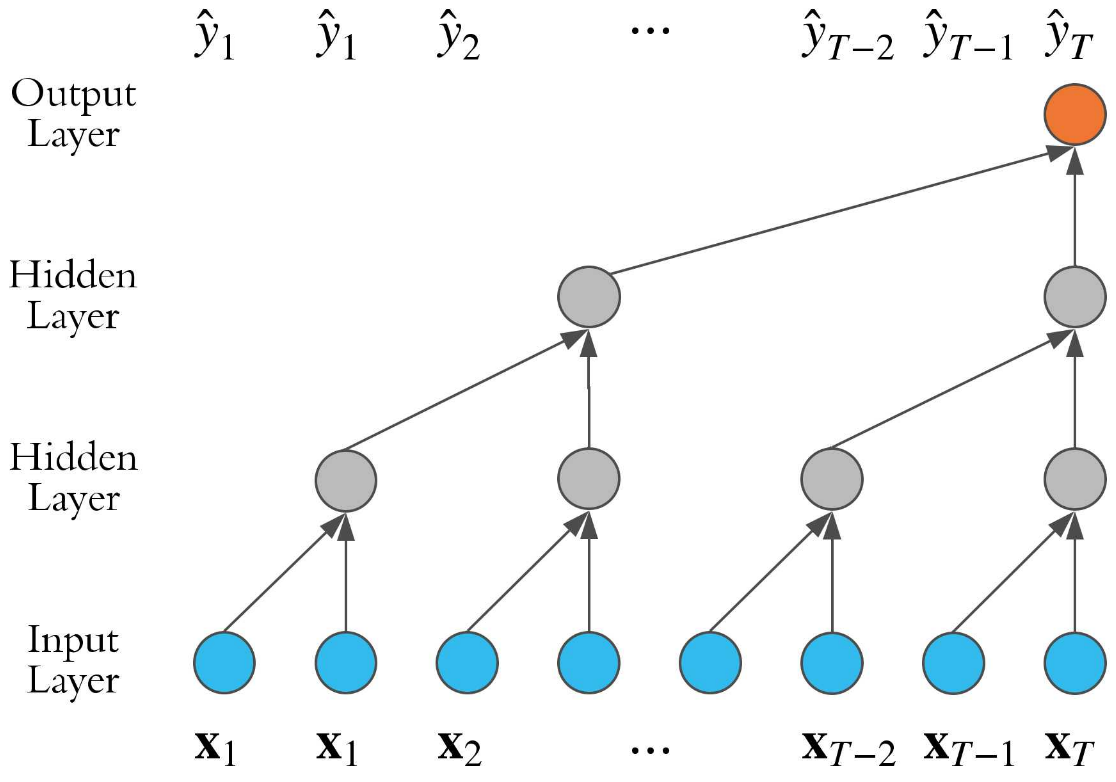
Causal Convolutions



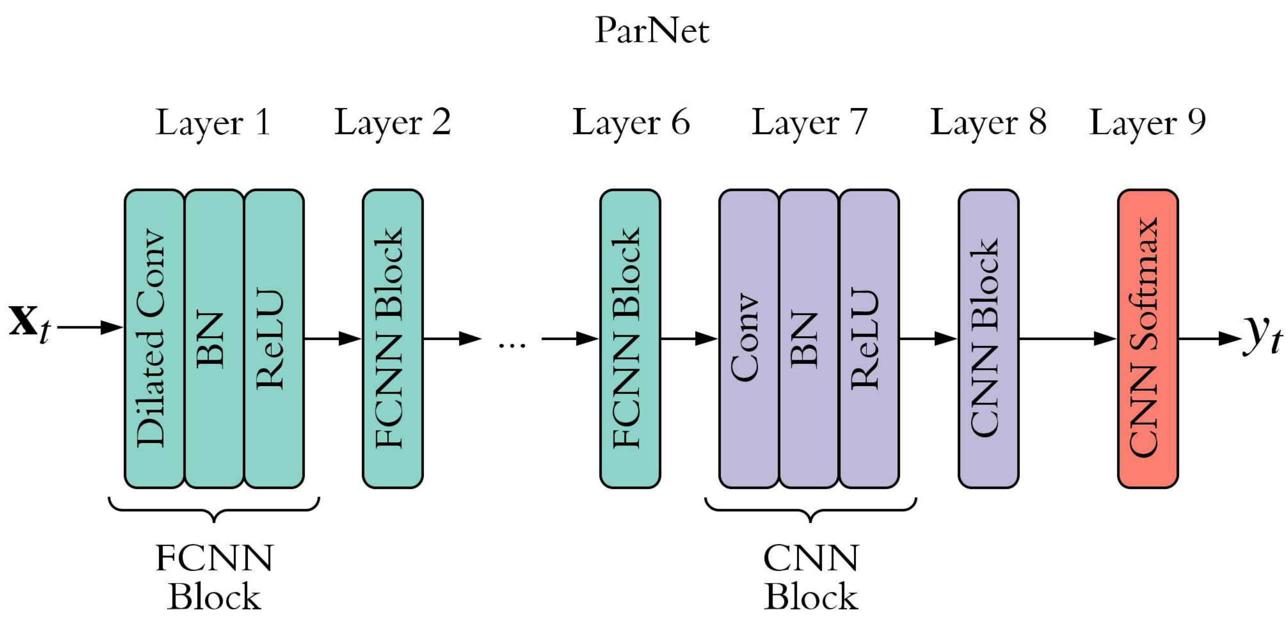
Causal Convolutions



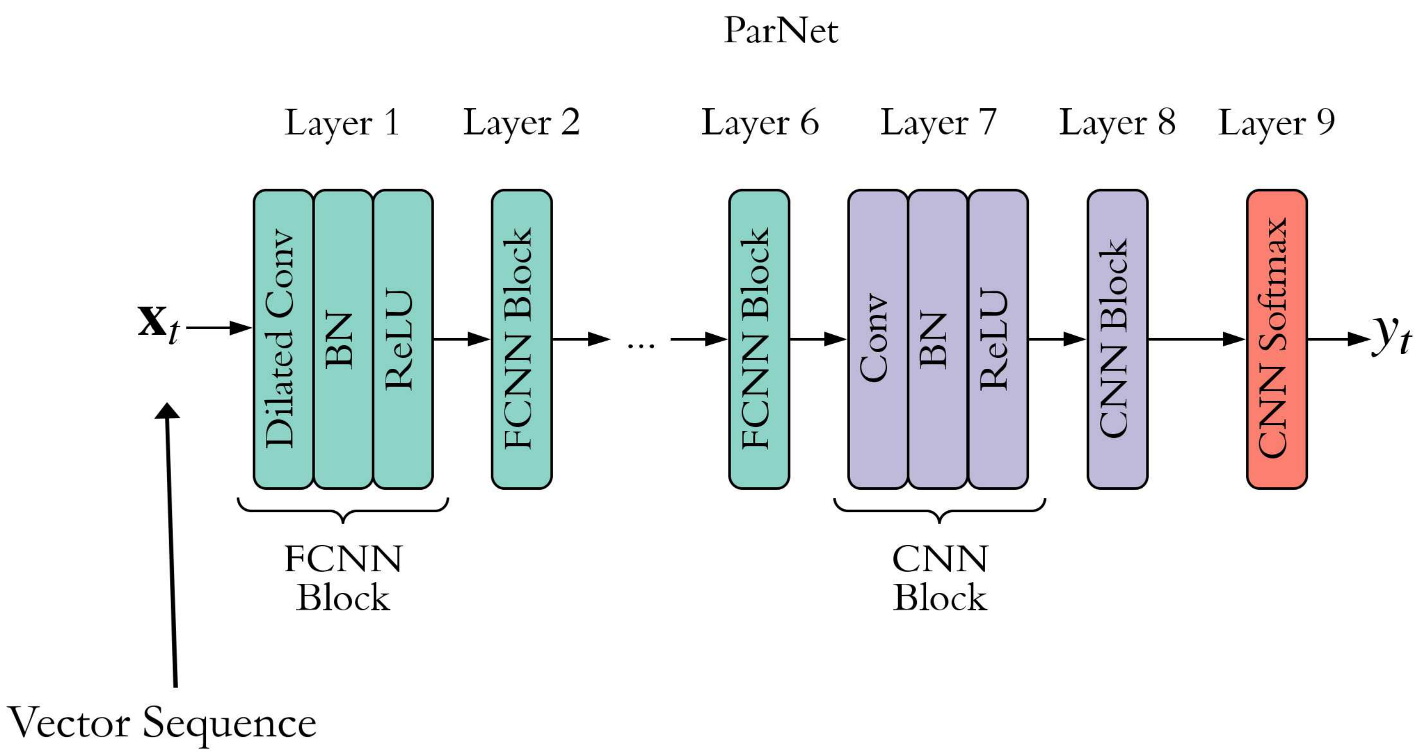
Causal Dilated Convolutions



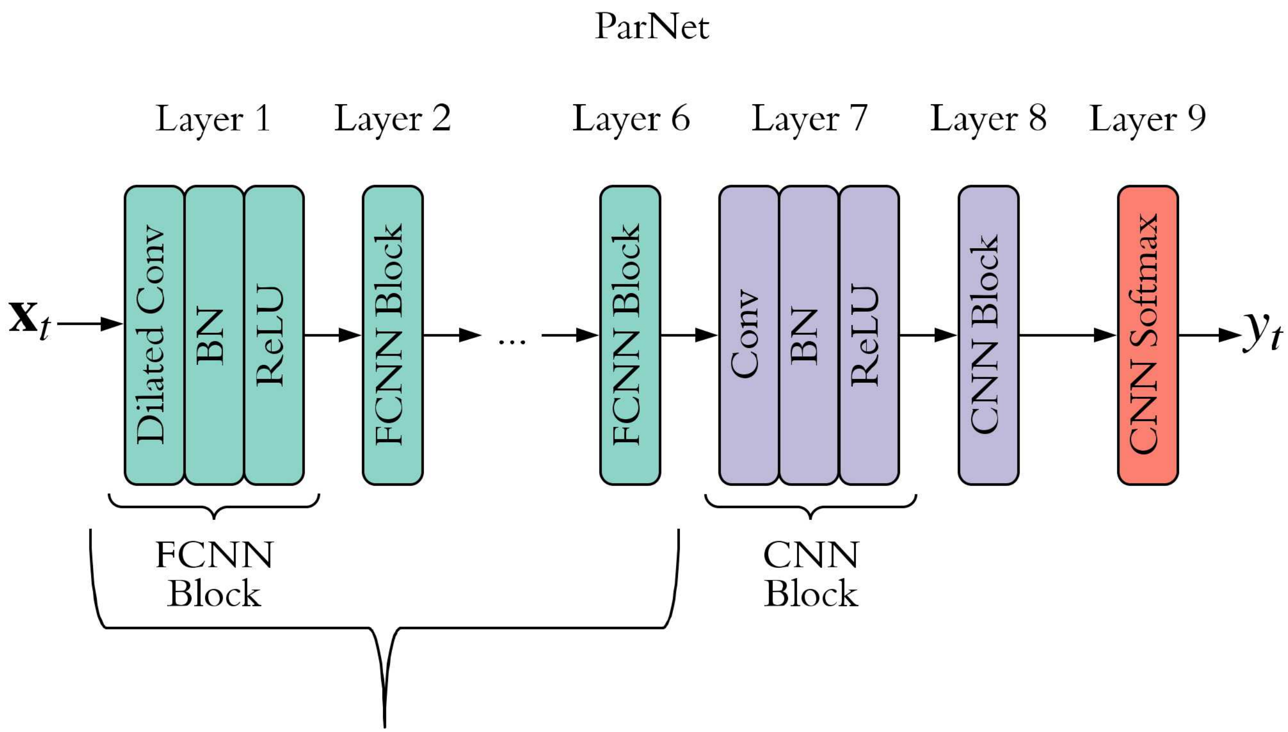
ParNet: Pedestrian Activity Network



ParNet: Pedestrian Activity Network



ParNet: Pedestrian Activity Network

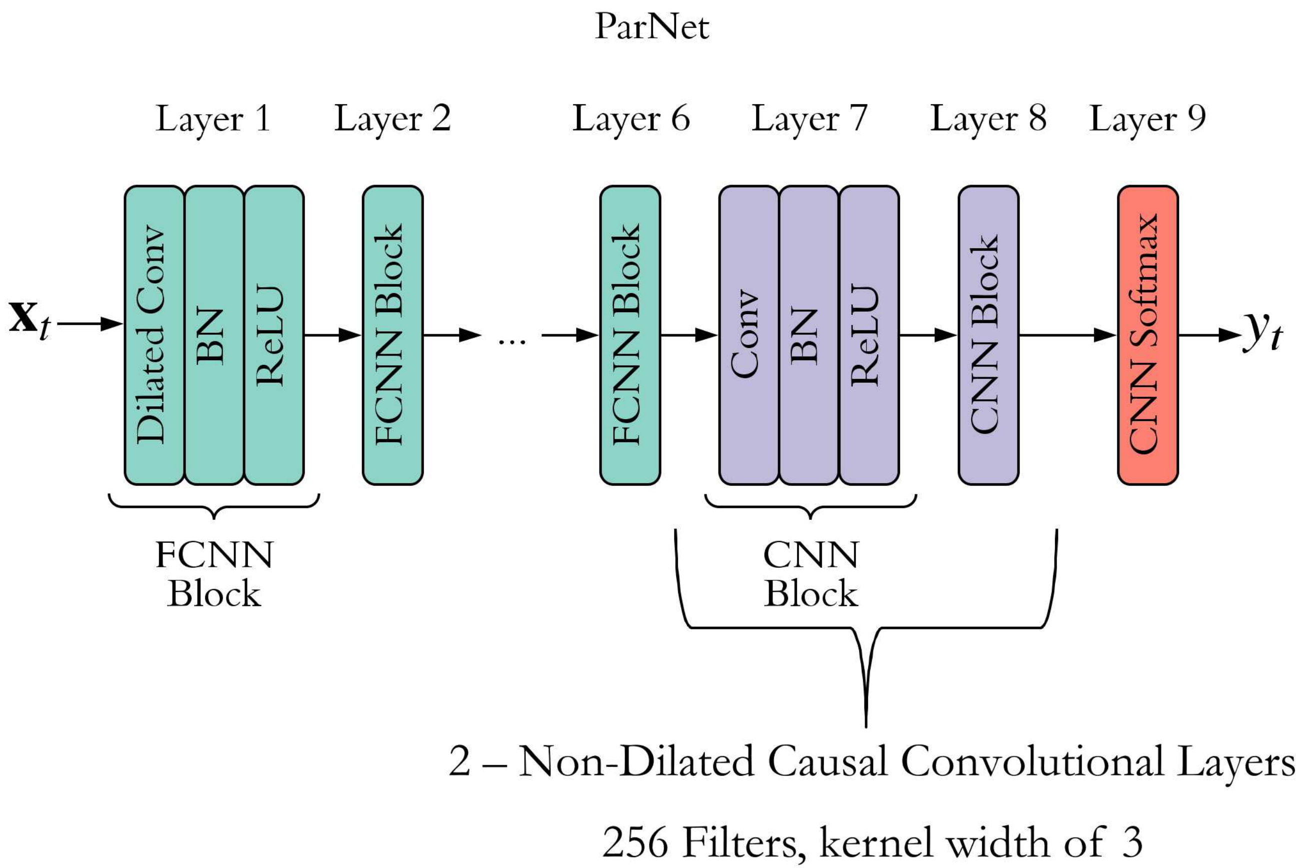


6 – Dilated Causal Convolutional Layers

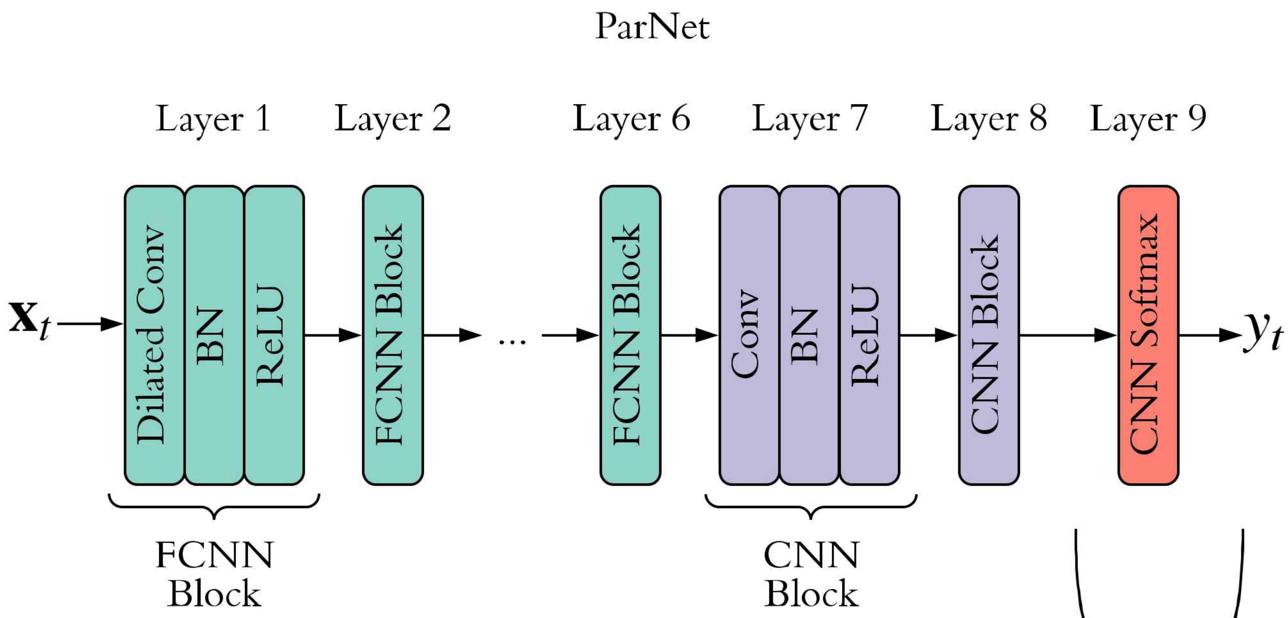
128 Filters, kernel width of 5

Dilation rates, 1, 2, 4, 8, 16, 32

ParNet: Pedestrian Activity Network



ParNet: Pedestrian Activity Network

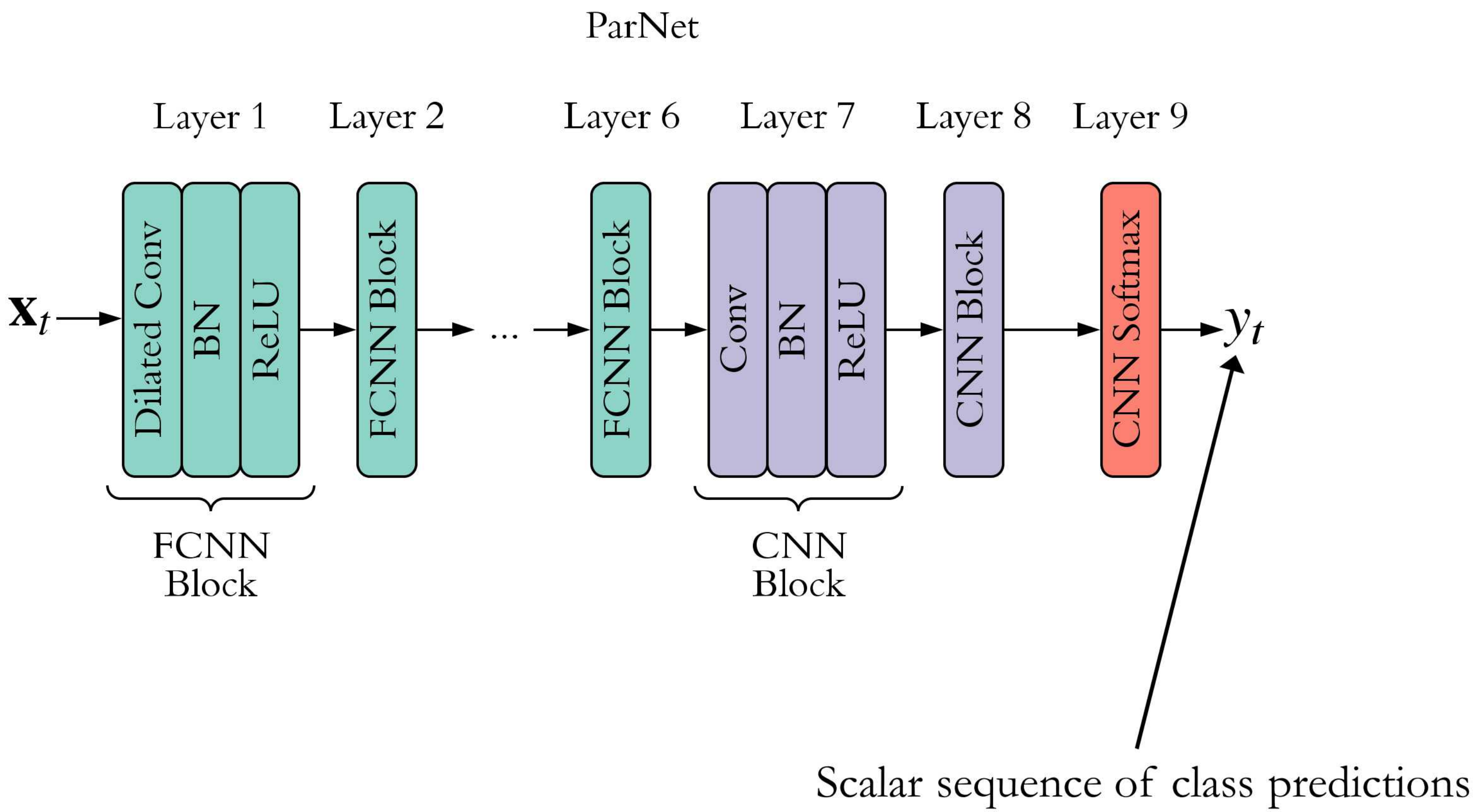


CNN with Softmax output

Kernel with of 1

Produces scaler sequence

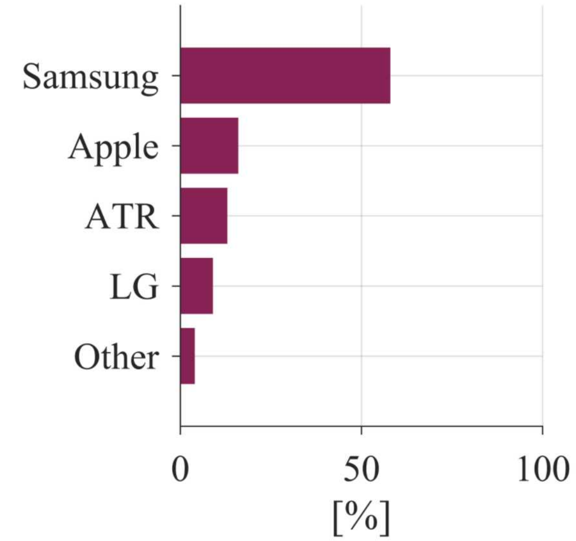
ParNet: Pedestrian Activity Network



Human Activity Sensing Consortium Dataset

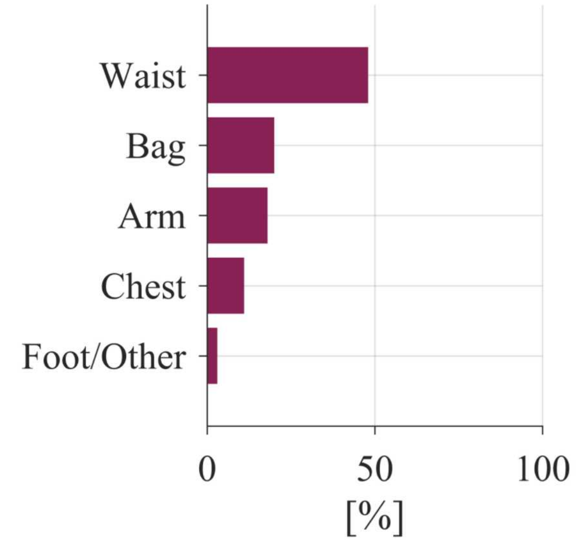
Data Sets for Human Activity Recognition

- Opportunity
- Physical Activity Monitoring Data Set
- UCI-Human Activity and Postural Transition Dataset
- Wireless Sensor Data Mining Dataset
- Human Activity Sensing Consortium



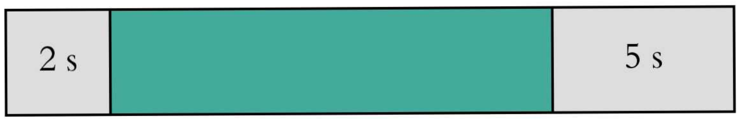
HASC-PAC2016

- 510 Participants (120 Femal, 390 Male)
- Segmented data (20 s duration)
- Data from accel/gyro/mag/barometer/proxy/wifi
- 111,027 segmented examples (for all sensor types)
- Activities: Stay, walk, run, skip, upstairs, downstairs
- Only used 100 Hz data
- 23,345 examples from all placement locations
- 5,970 examples from waist only placement



Human Activity Sensing Consortium Dataset

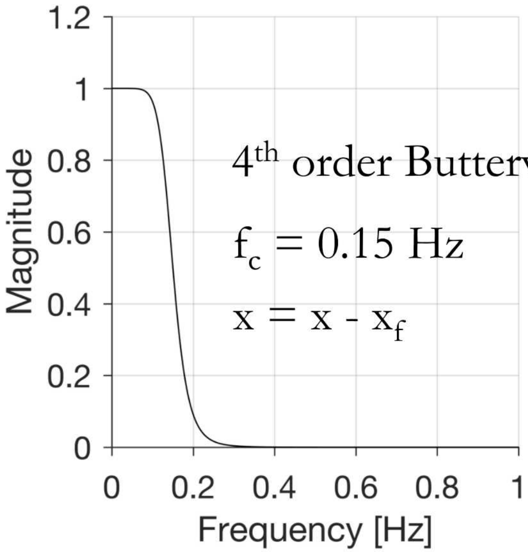
1 Crop Signal



Human Activity Sensing Consortium Dataset

1 Crop Signal

2 Remove Effects of Gravity



4^{th} order Butterworth

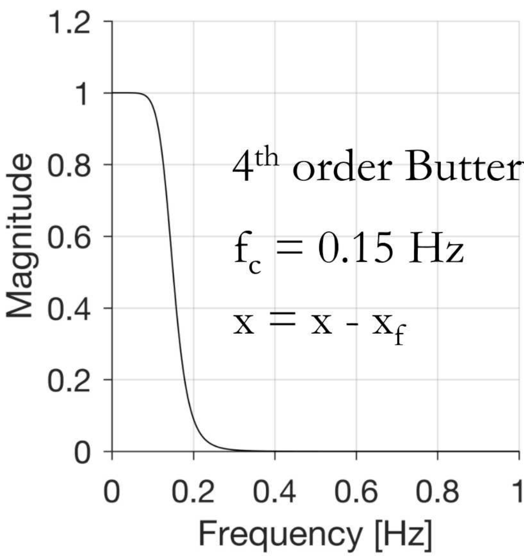
$f_c = 0.15$ Hz

$x = x - x_f$

Human Activity Sensing Consortium Dataset

1 Crop Signal

2 Remove Effects of Gravity



4th order Butterworth

$f_c = 0.15$ Hz

$x = x - x_f$

3 Noise Reduction

Digital polynomial smoothing filter

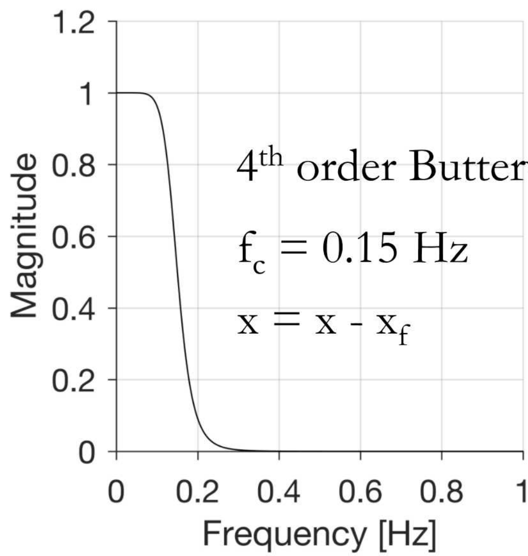
Savitzky-Golay filtering

Window = 51 and Order = 3

Human Activity Sensing Consortium Dataset

1 Crop Signal

2 Remove Effects of Gravity



4^{th} order Butterworth

$f_c = 0.15$ Hz

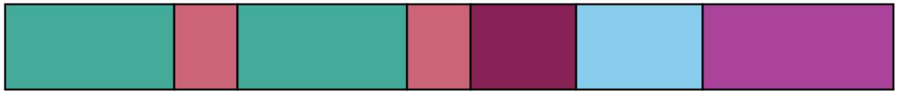
$x = x - x_f$

3 Noise Reduction

Digital polynomial smoothing filter
Savitzky-Golay filtering
Window = 51 and Order = 3

4 Data Augmentation

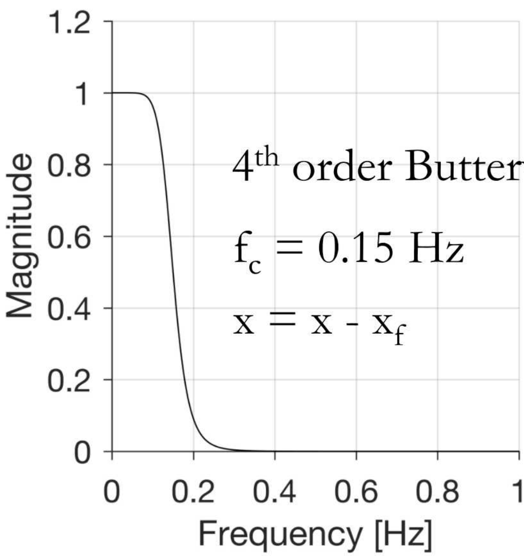
a Randomly Crop Signal



Human Activity Sensing Consortium Dataset

1 Crop Signal

2 Remove Effects of Gravity



4^{th} order Butterworth

$f_c = 0.15$ Hz

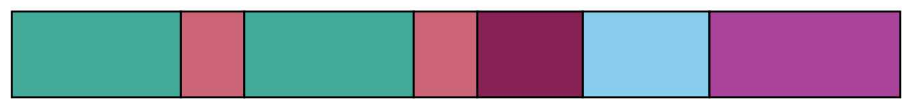
$x = x - x_f$

3 Noise Reduction

Digital polynomial smoothing filter
Savitzky-Golay filtering
Window = 51 and Order = 3

4 Data Augmentation

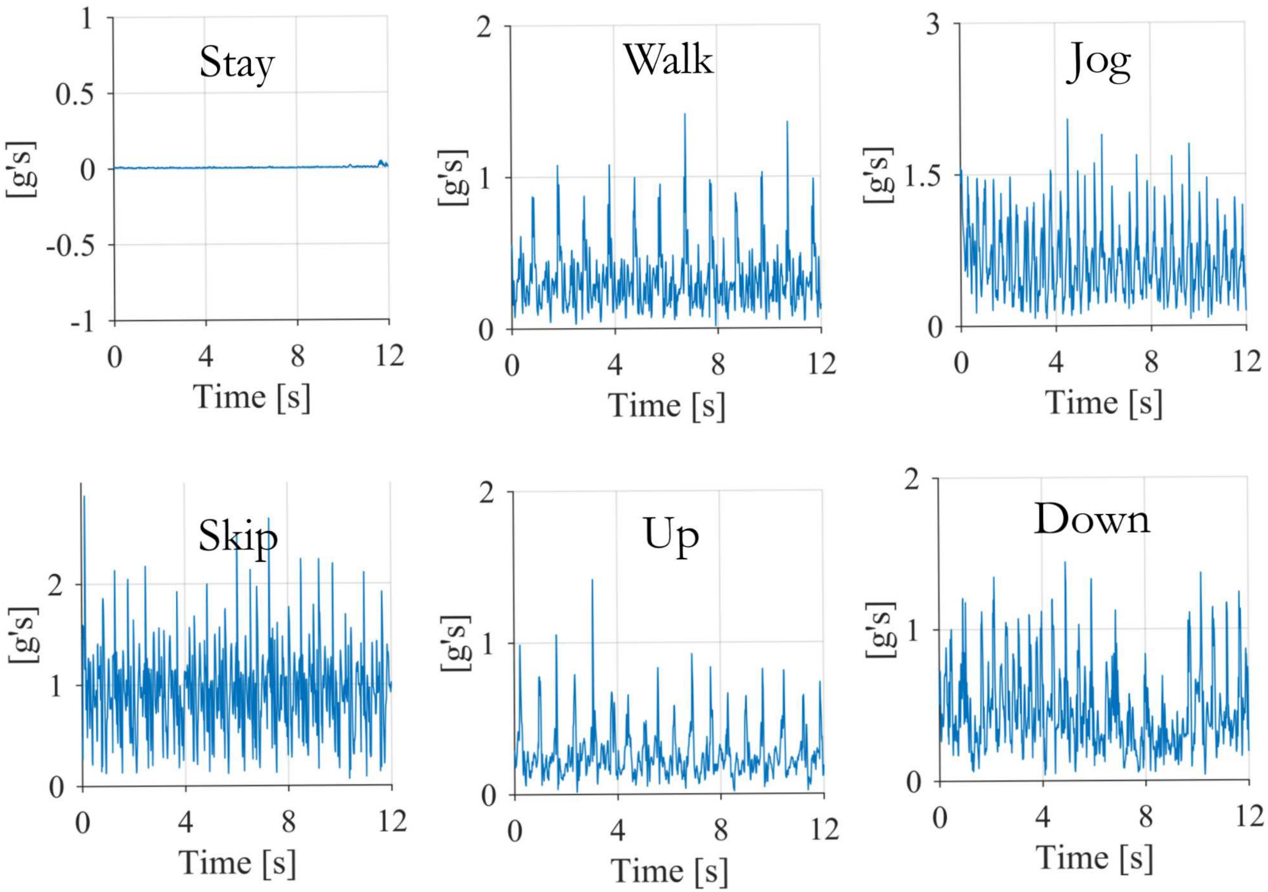
a Randomly Crop Signal



b Randomly Rotate Signals

Each axis is randomly rotated by 0° or 90°

Human Activity Sensing Consortium Dataset



Human Activity Sensing Consortium Data Set

Model	Placement	Sensor Type(s)	Examples
ParNet(All, Accel)	All	Accel	42,125
ParNet(All, Accel + Gyro)	All	Accel + Gyro	42,125
ParNet(Waist, Accel)	Waist	Accel	12,882
ParNet(Waist, Accel + Gyro)	Waist	Accel + Gyro	12,882

Pedestrian Activity Recognition: Training

Models trained using PyTorch

Trained and evaluated on 2x Nvidia GeForce® Gtx 980 GPUs

Each model was trained fully supervised for 250 epochs

Network Parameter Optimization

- Mini-batch gradient descent (batch size of 64 examples)
- minimize cross-entropy loss (measure of difference between probability distributions)
- Adaptive Moment Estimation (Adam) optimizer
- Learning rate of 1e-5
- L^2 regularization with coefficient of 10e-2

Pedestrian Activity Recognition: Training

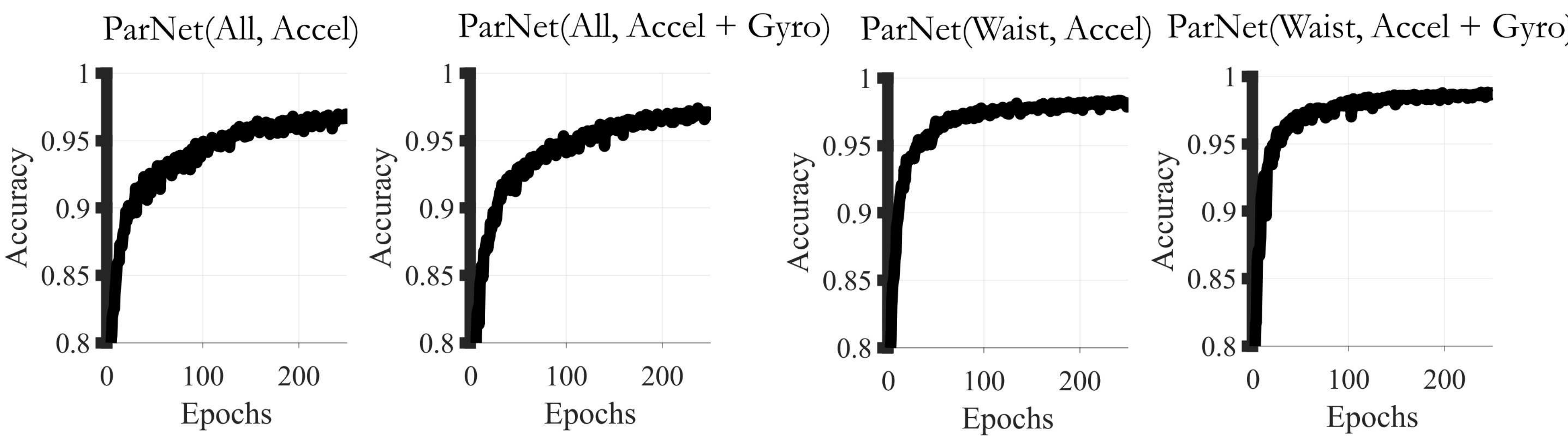
Models trained using PyTorch

Trained and evaluated on 2x Nvidia GeForce® Gtx 980 GPUs

Each model was trained fully supervised for 250 epochs

Network Parameter Optimization

- Mini-batch gradient descent (batch size of 64 examples)
- minimize cross-entropy loss (measure of difference between probability distributions)
- Adaptive Moment Estimation (Adam) optimizer
- Learning rate of 1e-5
- L^2 regularization with coefficient of 10e-2



ParNet models with accel + gyro have better performance than accel only

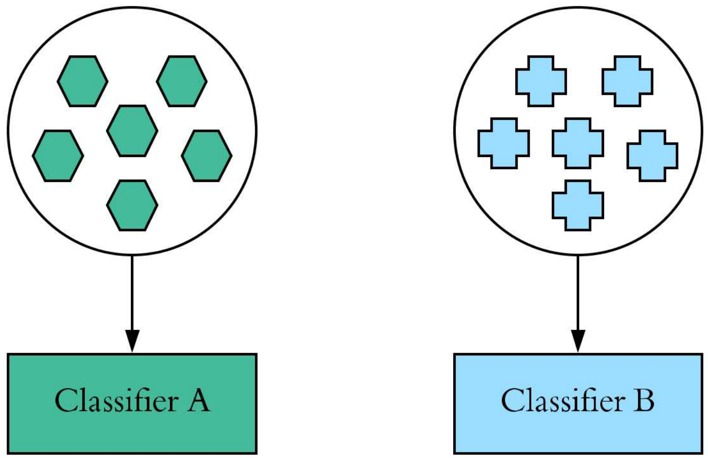
Pedestrian Activity Recognition: Evaluation

Model	Acc	Stay	Walk	Jog	Skip	Up	Down
ParNet(All, Accel)	97.0%	99.6%	95.6%	97.4%	98.2%	98.0%	93.4%
ParNet(All, Accel+Gyro)	**97.4%**	**99.8%**	**96.5%**	**98.0%**	**98.4%**	**97.7%**	**94.5%**
ParNet(Waist, Accel)	98.3%	99.7%	97.7%	98.7%	97.6%	97.6%	98.7%
ParNet(Waist, Accel + Gyro)	**98.8%**	**99.4%**	**98.3%**	**99.7%**	**98.5%**	**98.7%**	**98.4%**
Random Forests(All, Accel)	73.4%	-	-	-	-	-	-
Random Forests(All, Waist)	81.4%	-	-	-	-	-	-
RNN	95.4%	-	-	-	-	-	-

ParNet has equivalent or better performance than previously published results

Transfer Learning For Neural Networks

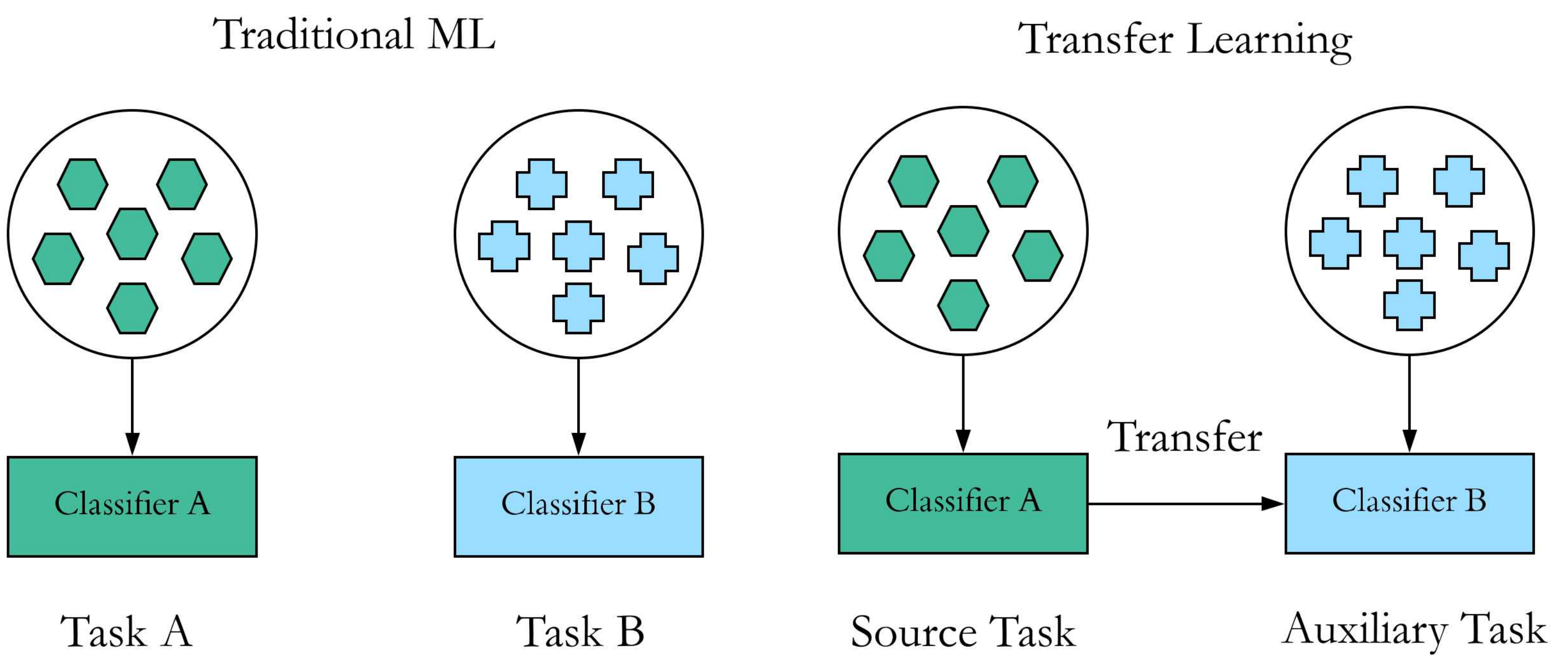
Traditional ML



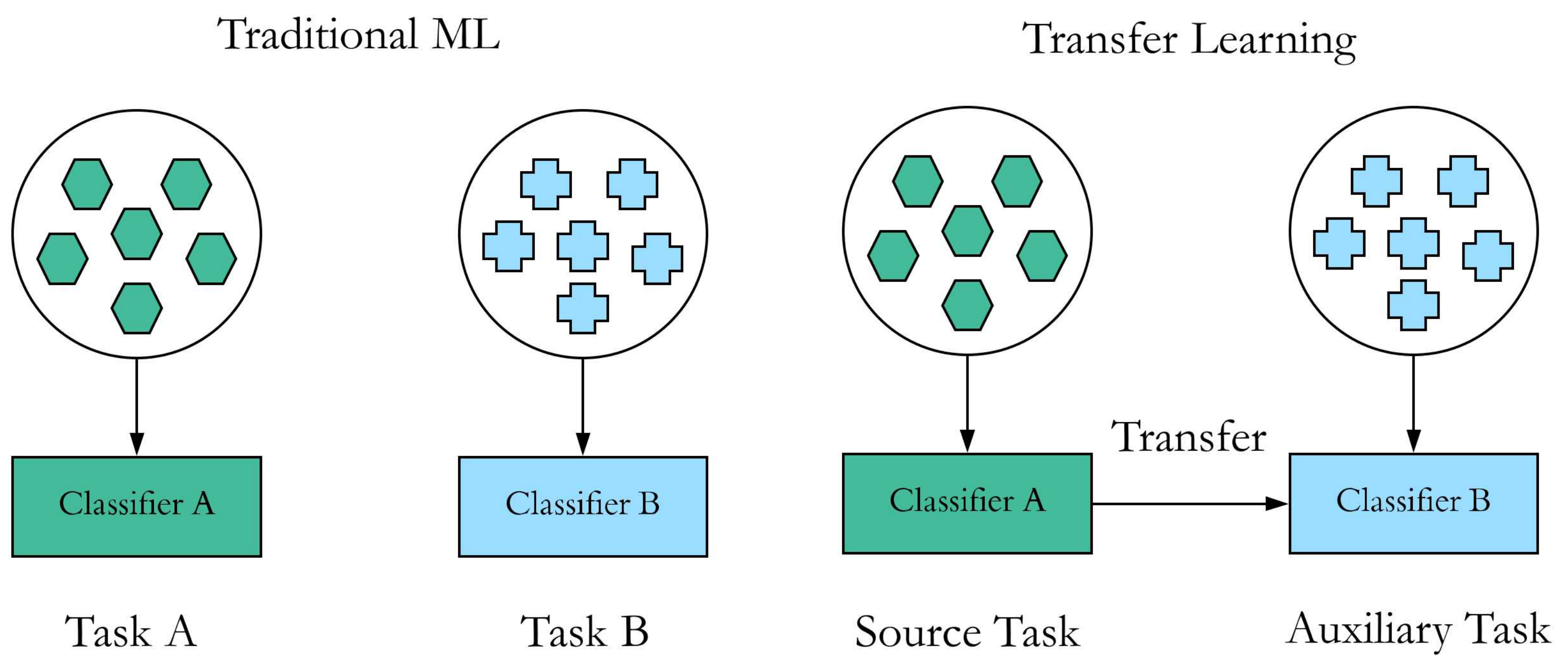
Task A

Task B

Transfer Learning For Neural Networks



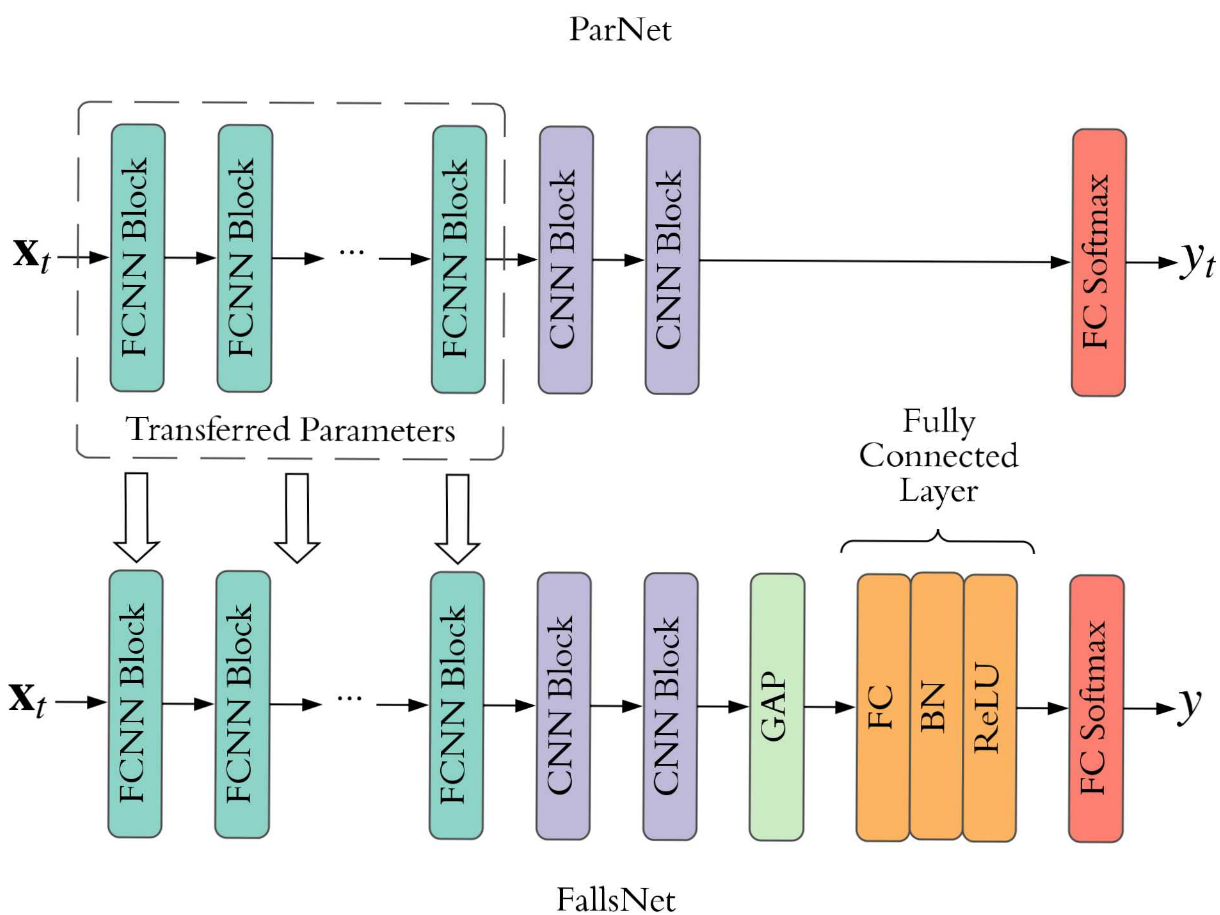
Transfer Learning For Neural Networks



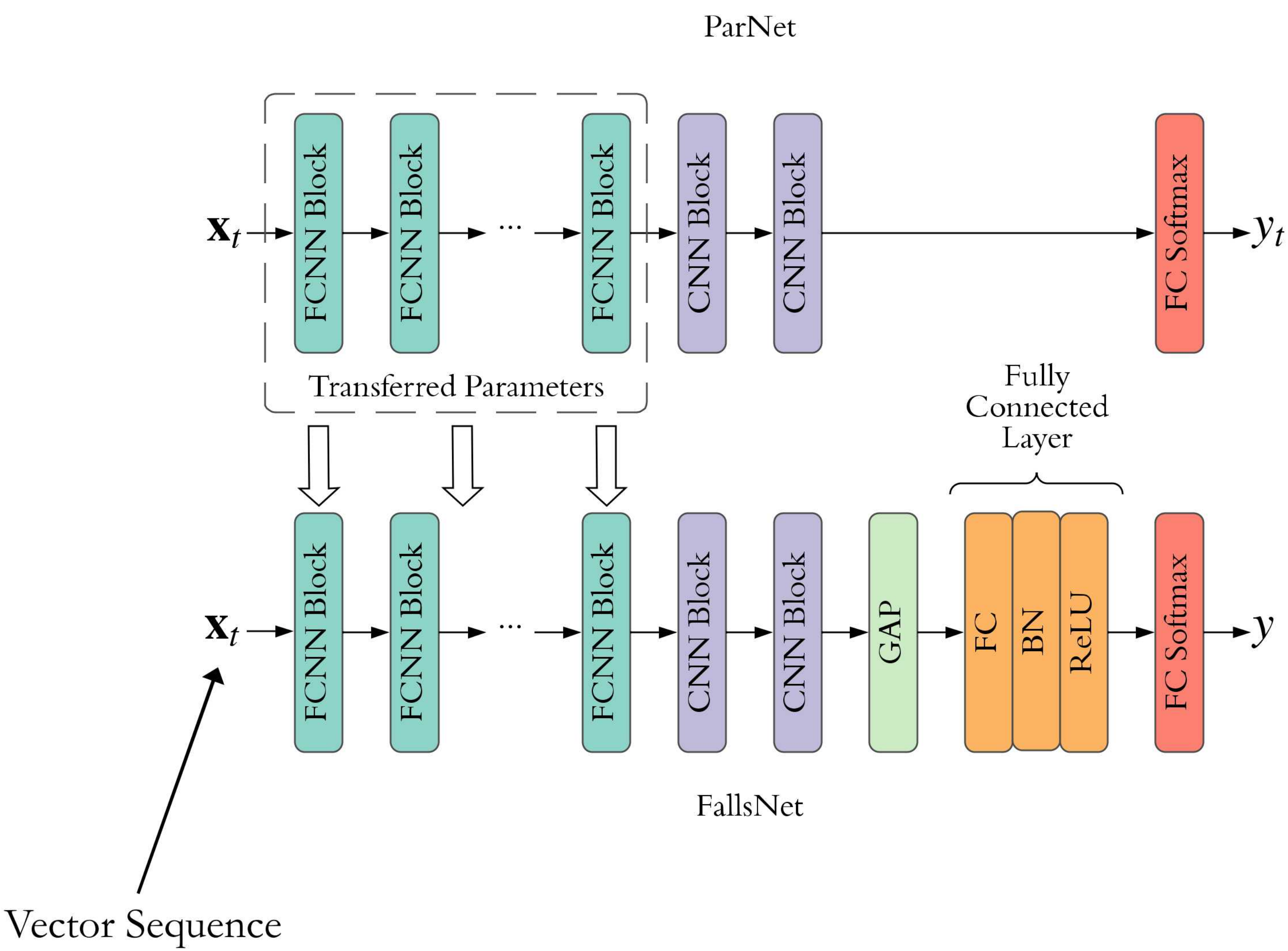
Transfer Learning for Neural Networks

- Transfer weights from classifier A to classifier B. Retrain only output layer.
- Transfer some weights from classifier A to classifier B. Retrain layers with non-transferred weights.
- Transfer weights from classifier A to classifier B. Set layer-wise learning rates with rates increasing with adepth. Retrain output layer

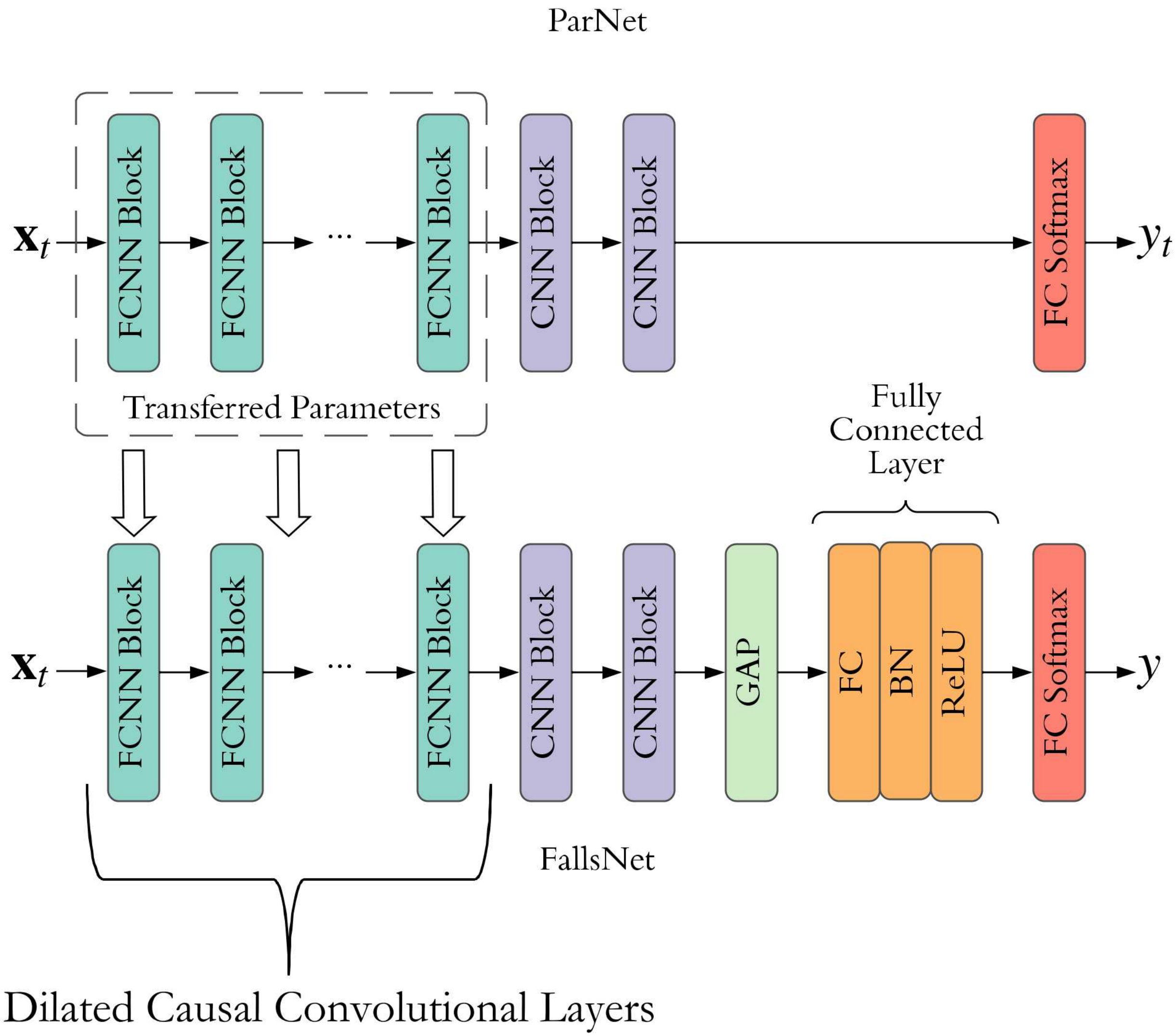
FallsNet: Falls Risk Classification Network



FallsNet: Falls Risk Classification Network

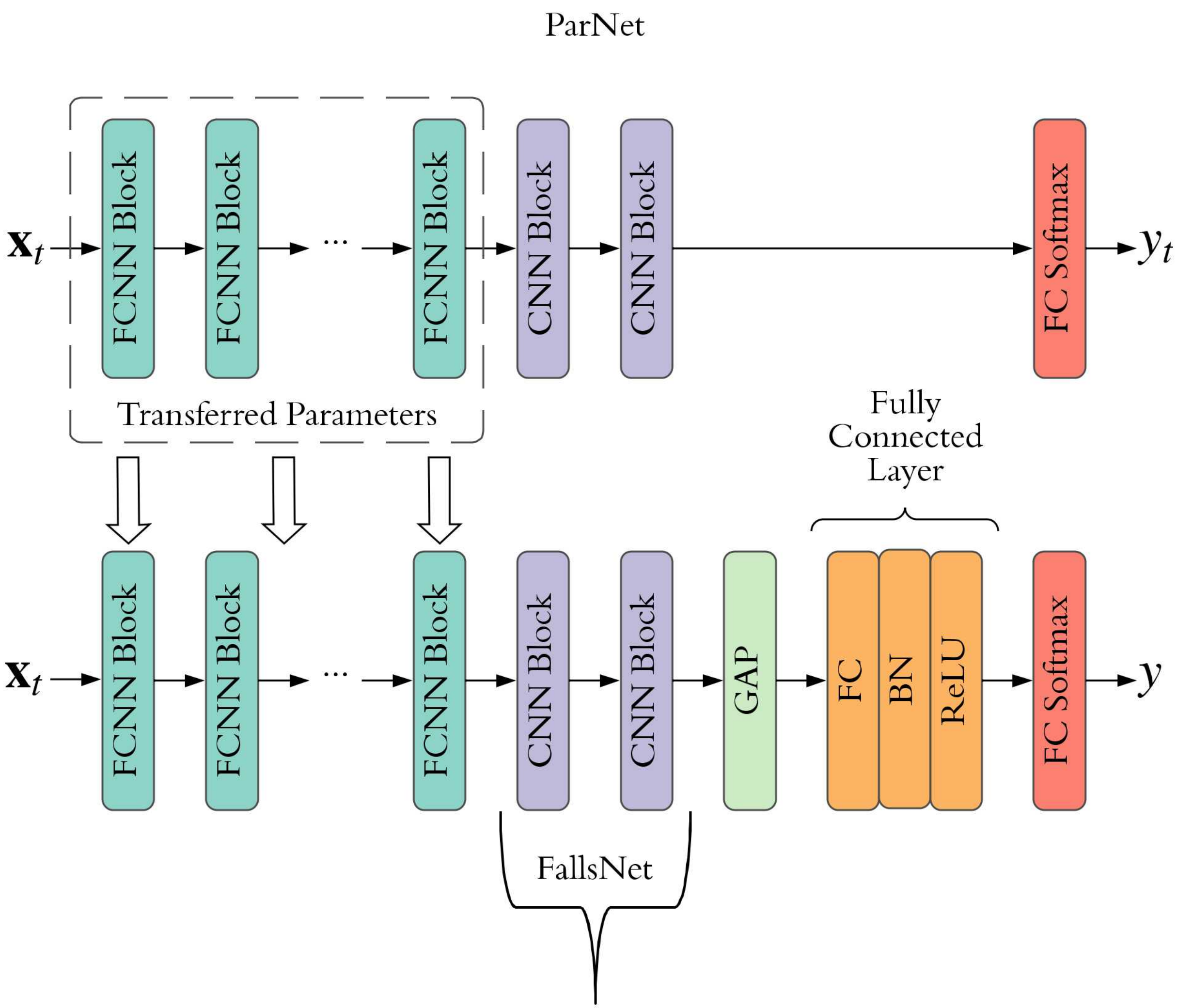


FallsNet: Falls Risk Classification Network



128 Filters, kernel width of 5, Dilation rates, 1, 2, 4, 8, 16, 32

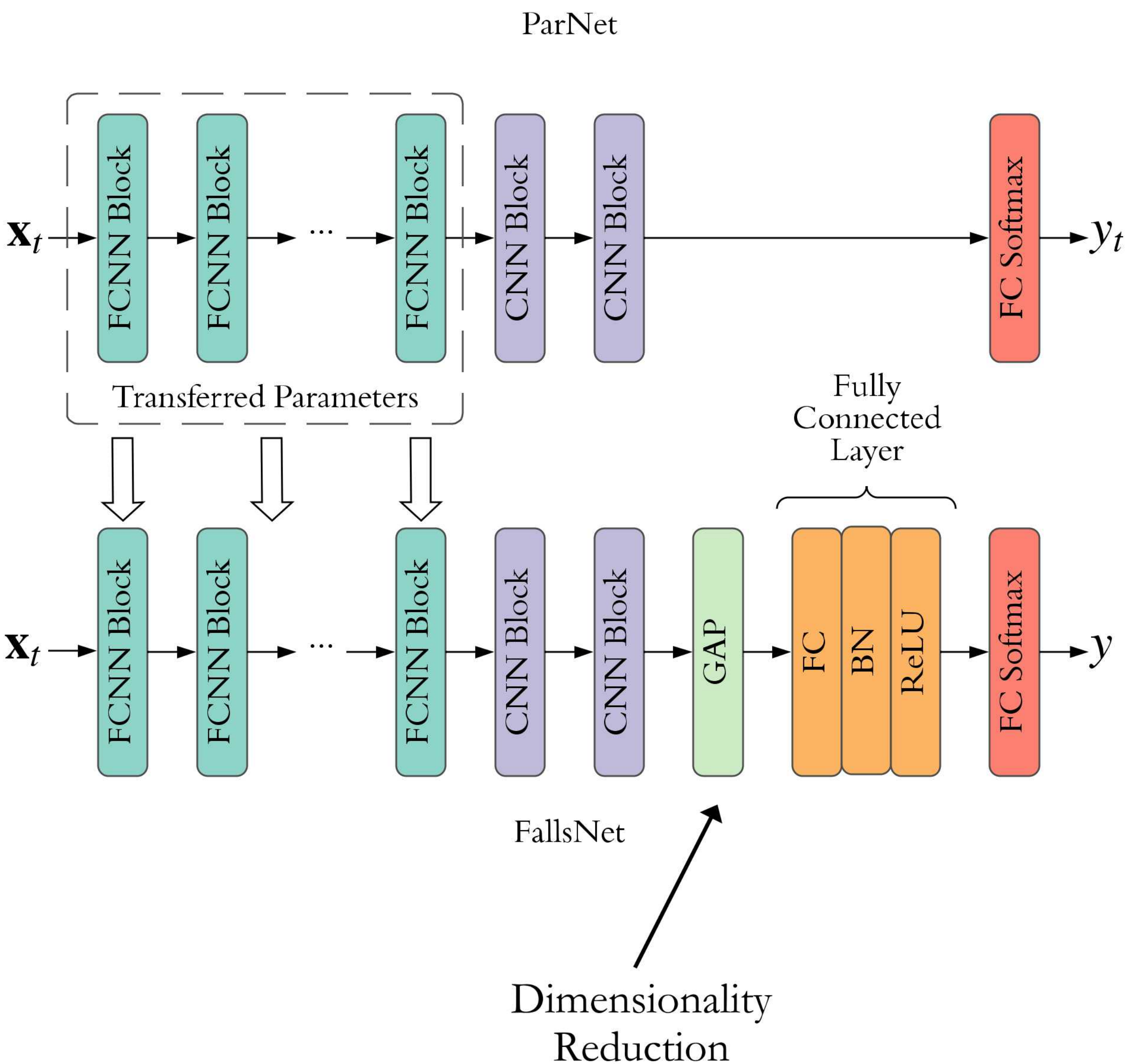
FallsNet: Falls Risk Classification Network



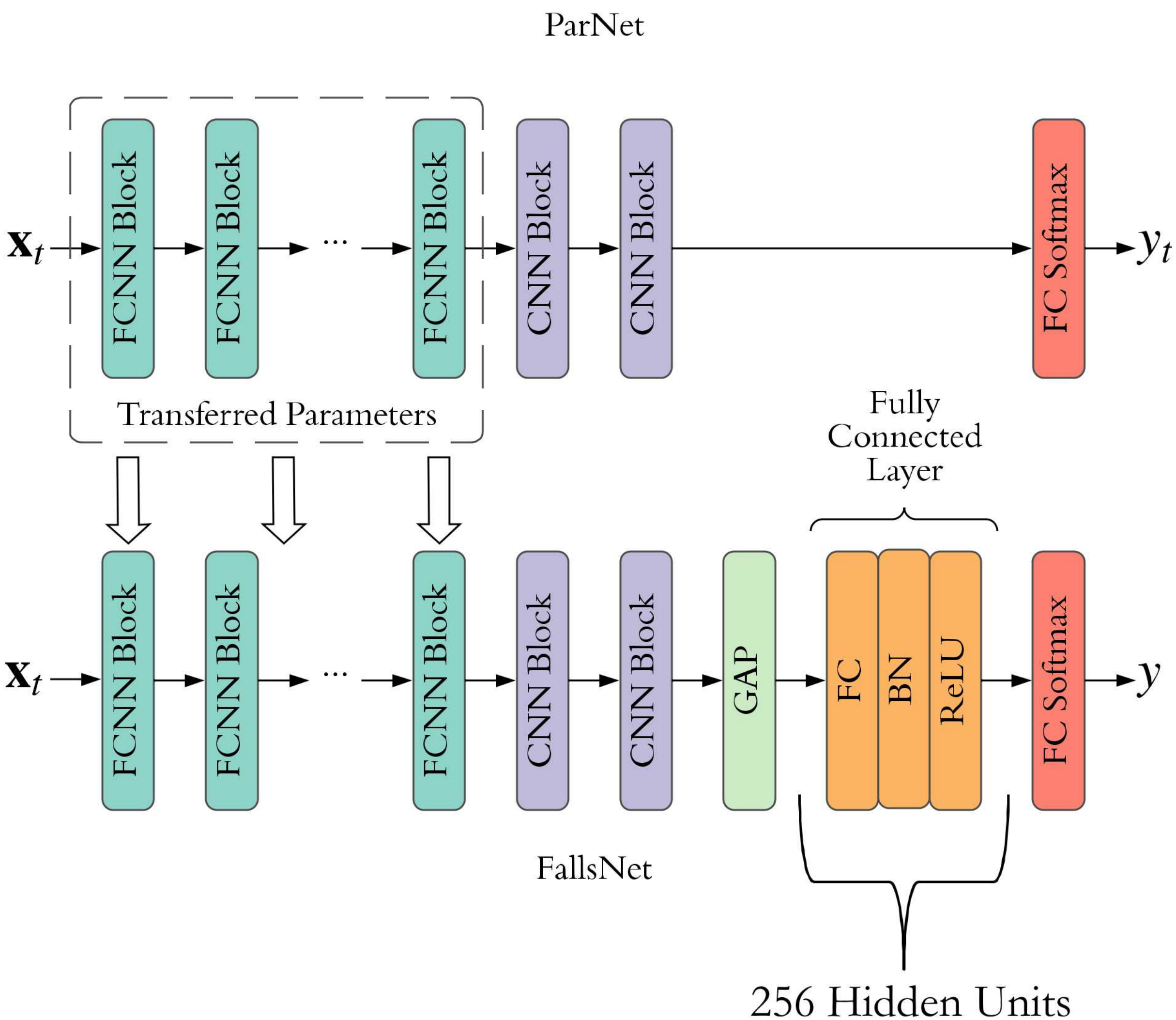
2 – Non-Dilated Causal Convolutional Layers

256 Filters, kernel width of 3

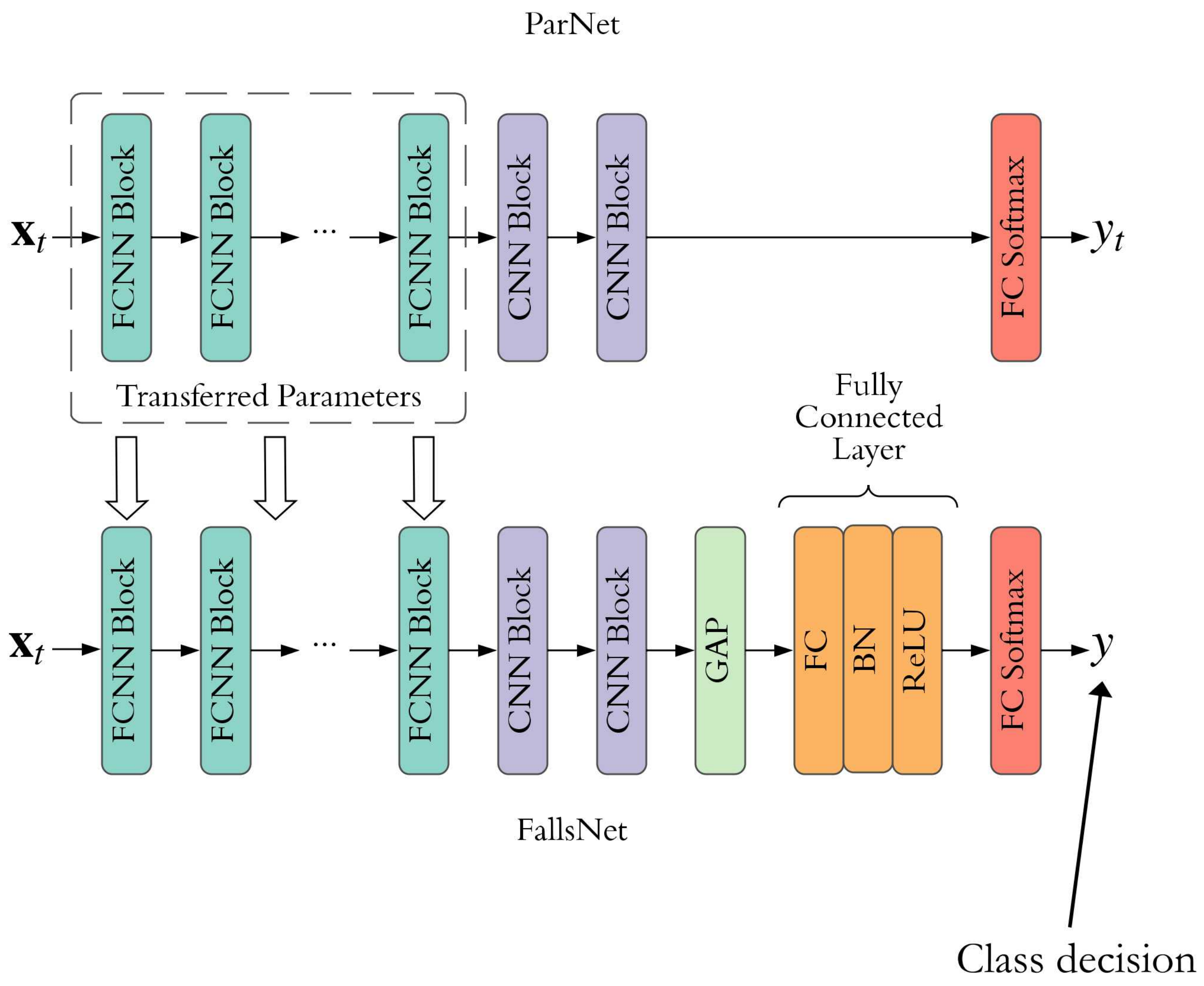
FallsNet: Falls Risk Classification Network



FallsNet: Falls Risk Classification Network



FallsNet: Falls Risk Classification Network



Older Adult Smartphone Data

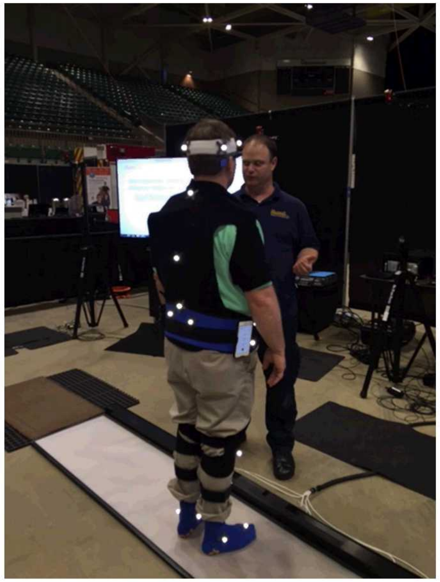
Data collected in partnership with the Electronic Caregiver Company

Sensor System:

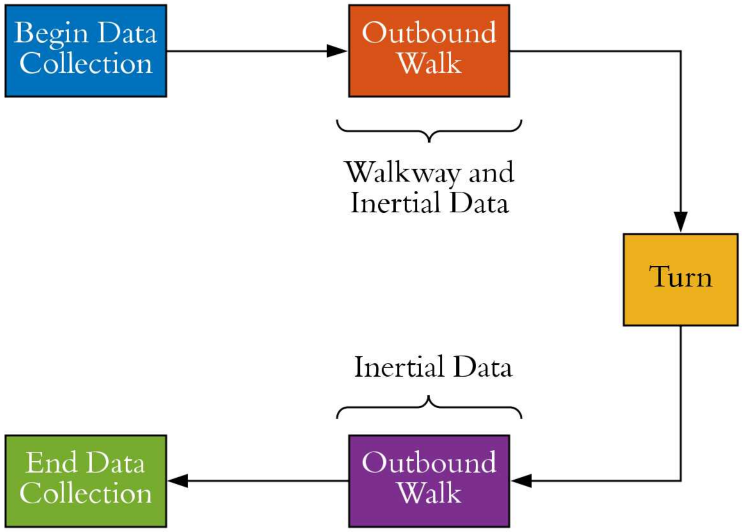
- TekScan® Walkway™ System
- 2x Apple iPhone 6 (custom iOS app)
- Inertial measurements of gait
- 6 sensor channels, 3-axis accel, 3-axis gyro

Smartphone Data Collection:

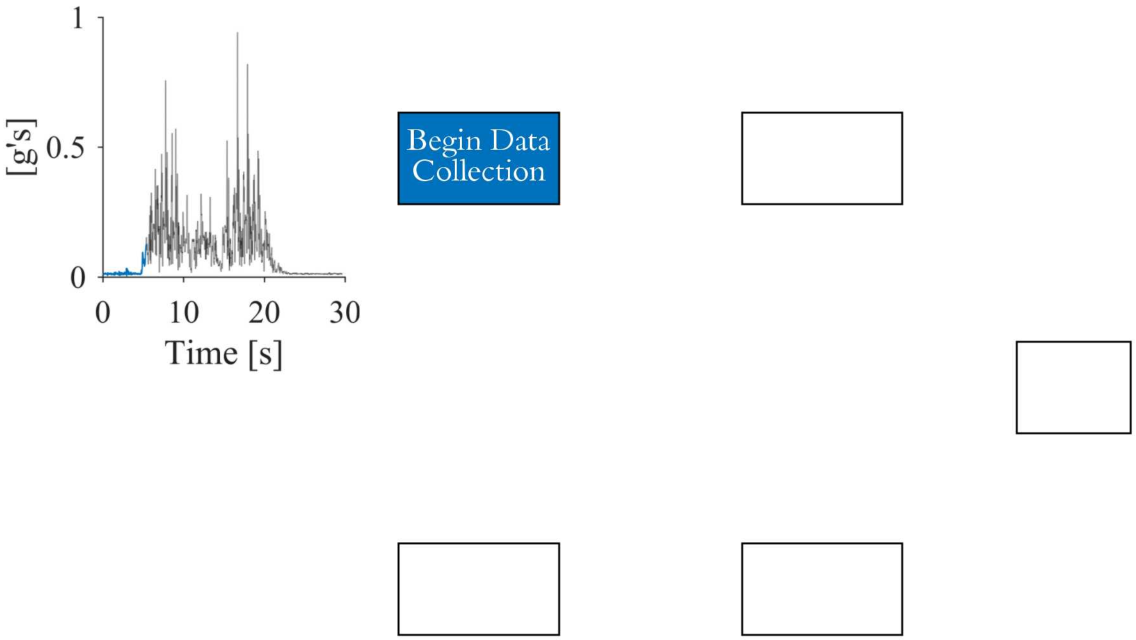
- Data collected from 256 of 854 participants
- Attached to left and right hip using holster clip and gait belt
- Each participant has an example of walkway (labeling) and inertial gait data (prediction)



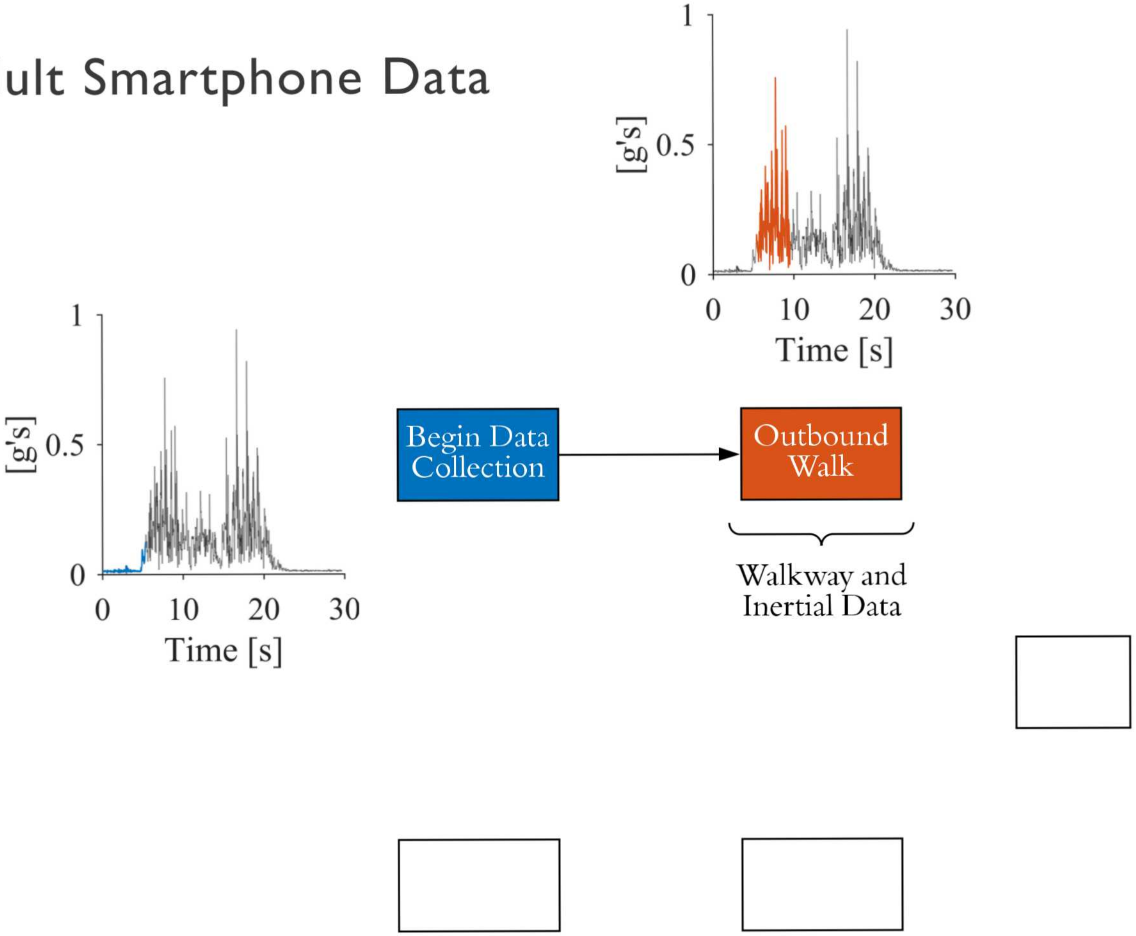
Older Adult Smartphone Data



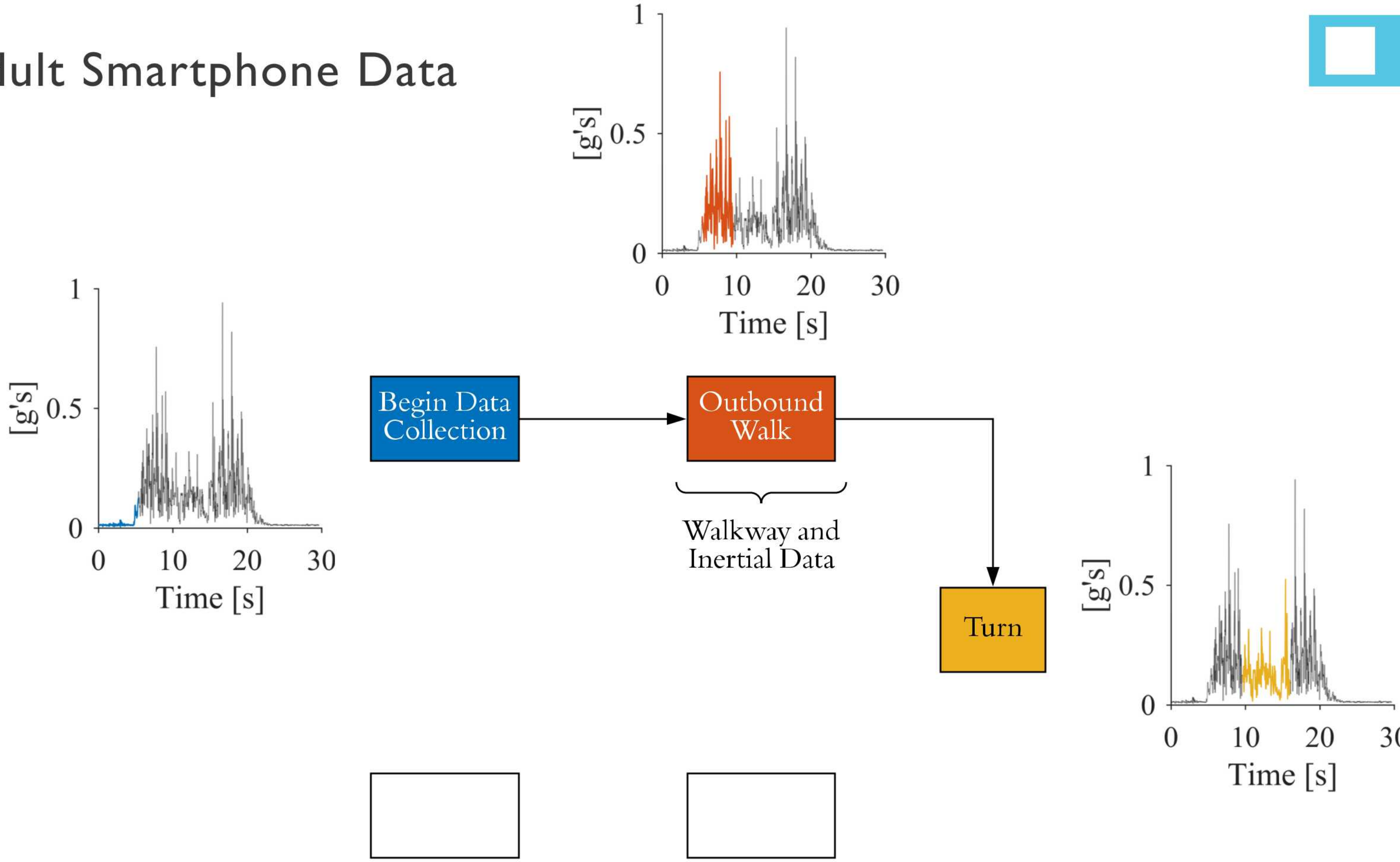
Older Adult Smartphone Data



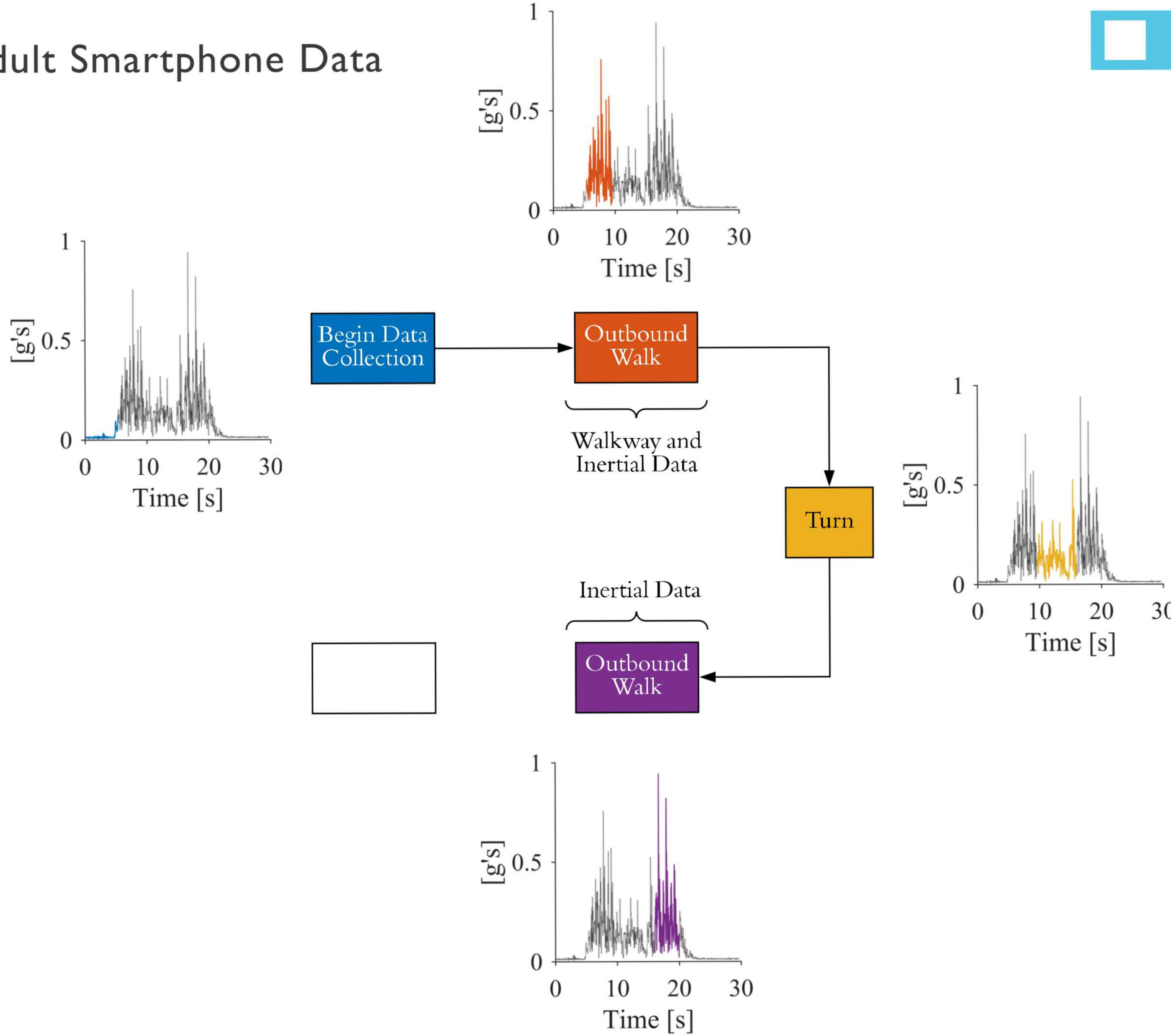
Older Adult Smartphone Data



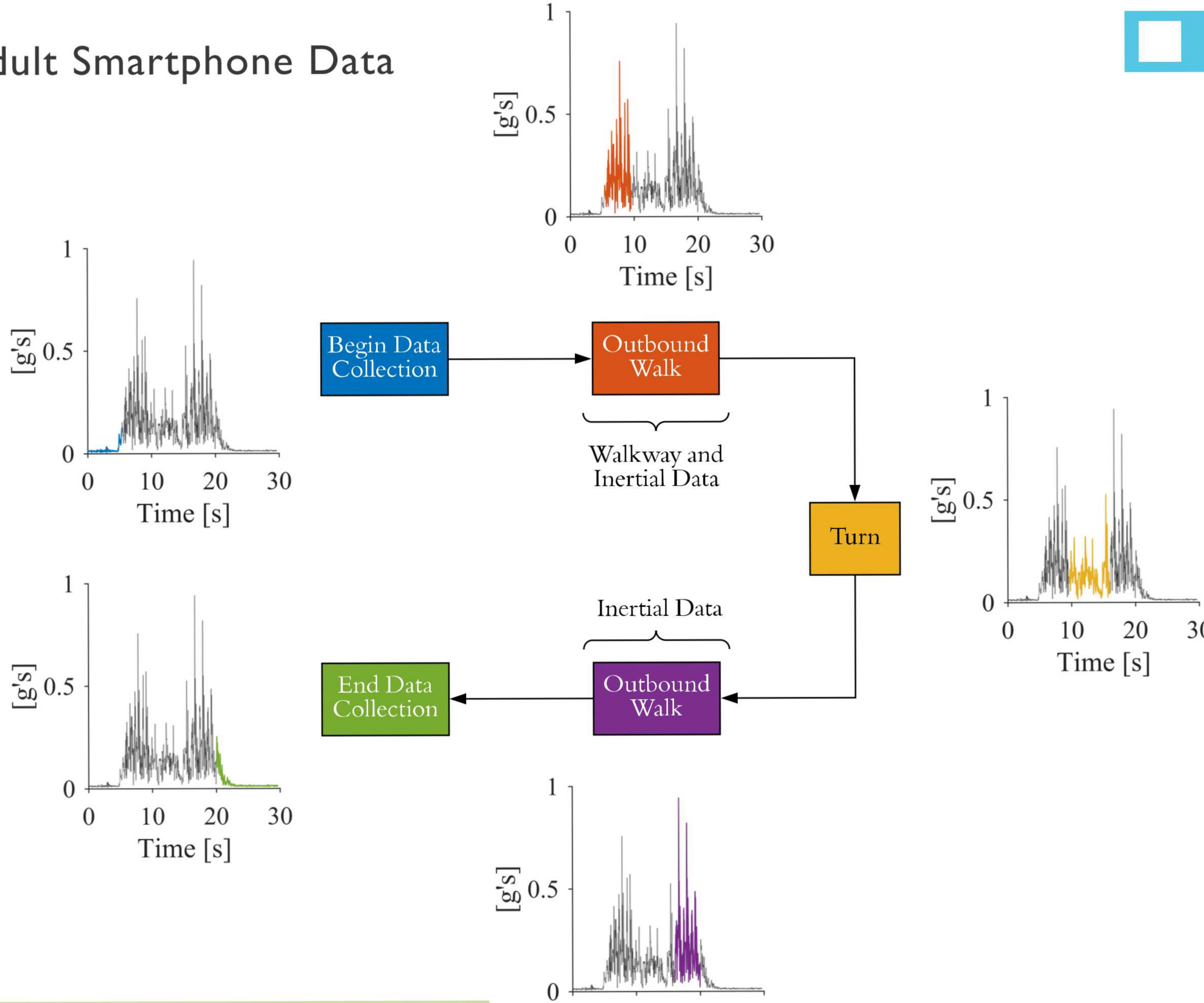
Older Adult Smartphone Data



Older Adult Smartphone Data



Older Adult Smartphone Data



Activities Are Manually Segmented

Older Adult Gait Data

Inertial Gait Examples

- 657 examples, where the duration is greater than 3.5 seconds
- 436 from female participants
- 221 from male participants

Gait Segment Labeling

- Each gait segment has a corresponding vector of gait variables
- Used Bayesian classifier to label each gait segment
- 422 examples labeled as low risk
- 235 examples labeled as high risk

Additional Post-Processing

- Digital polynomial smoothing filter
- Savitzky-Golay filtering
- Window = 51 and Order = 3

Transfer Learning Experiment

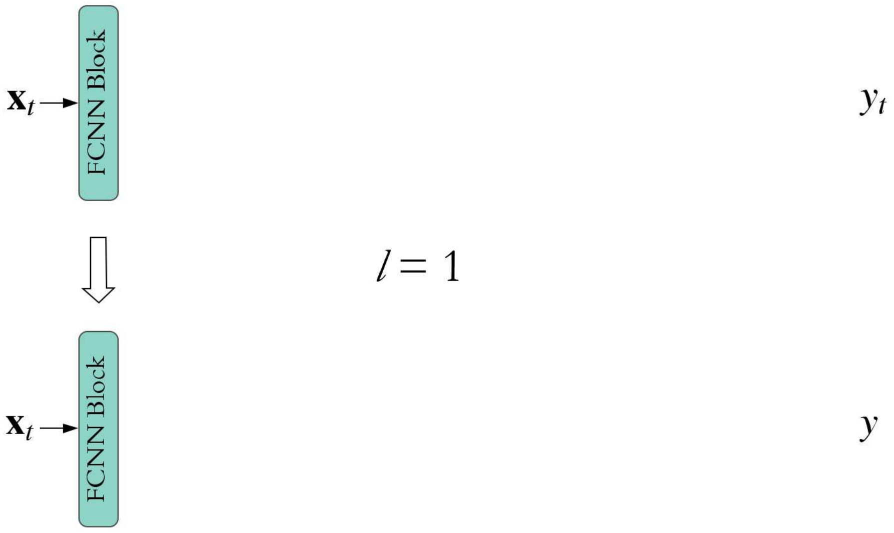
Transfer Learning Experiment

- First l layers of ParNet are transferred to the first l layers of FallsNet
- Remaining 6 – l layers are randomly initialized
- Only backpropagate through the randomly initialized layers
- Trained 24 different models

Transfer Learning Experiment

Transfer Learning Experiment

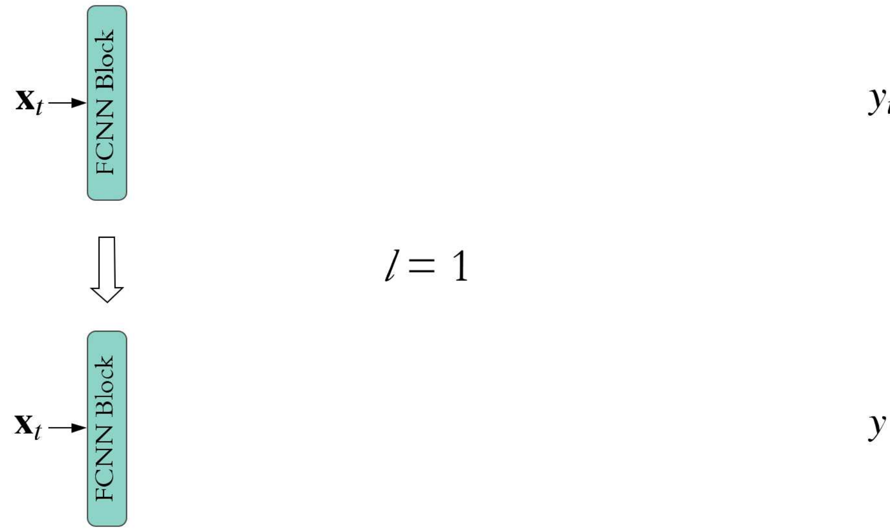
- First l layers of ParNet are transferred to the first l layers of FallsNet
- Remaining $6 - l$ layers are randomly initialized
- Only backpropagate through the randomly initialized layers
- Trained 24 different models

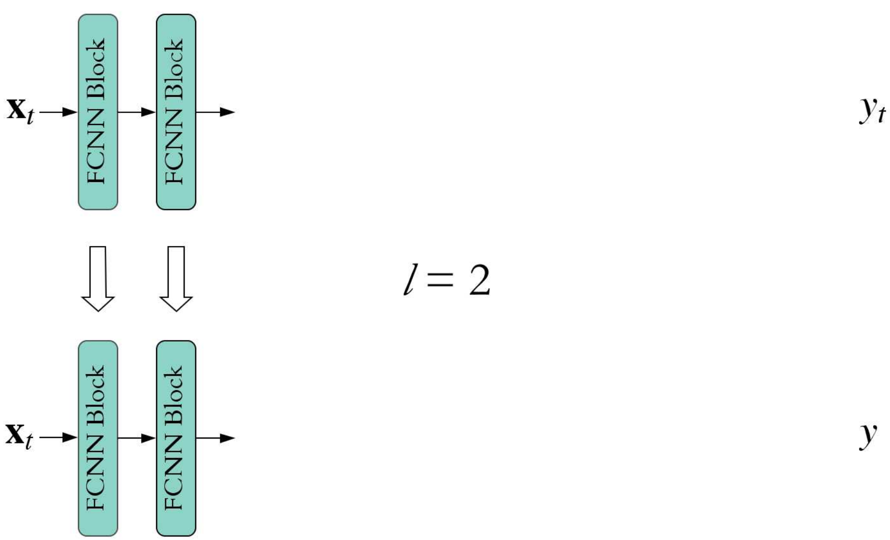


Transfer Learning Experiment

Transfer Learning Experiment

- First l layers of ParNet are transferred to the first l layers of FallsNet
- Remaining $6 - l$ layers are randomly initialized
- Only backpropagate through the randomly initialized layers
- Trained 24 different models

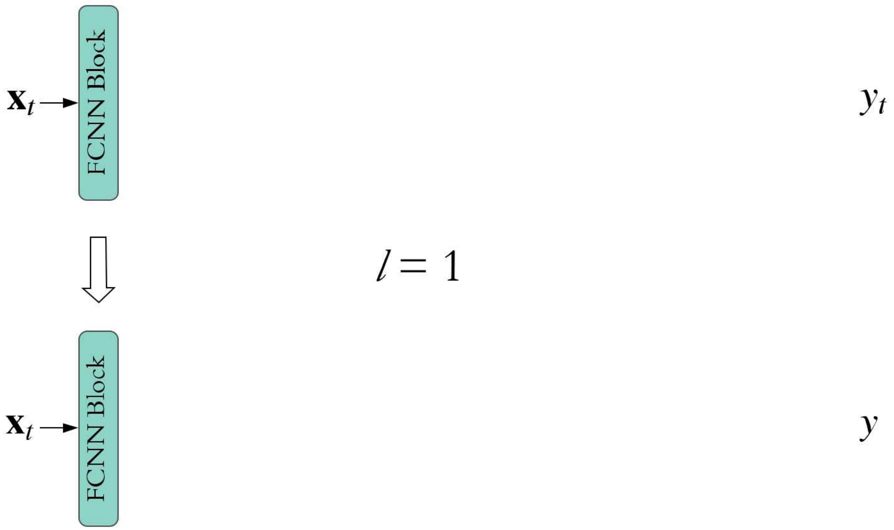


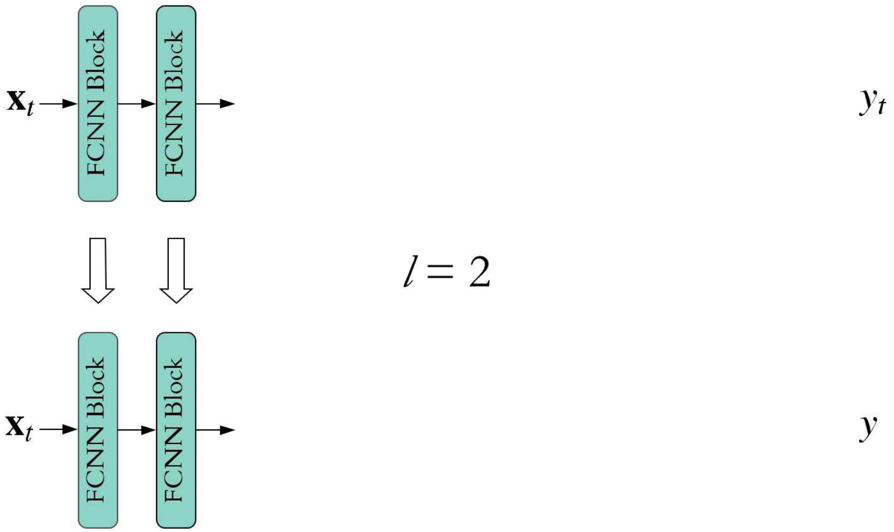


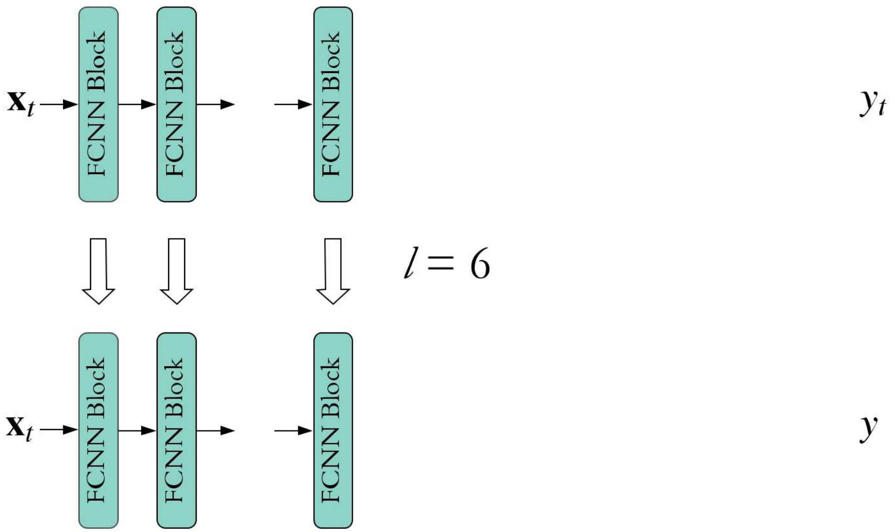
Transfer Learning Experiment

Transfer Learning Experiment

- First l layers of ParNet are transferred to the first l layers of FallsNet
- Remaining $6 - l$ layers are randomly initialized
- Only backpropagate through the randomly initialized layers
- Trained 24 different models
- Trained 2 baseline models ($l = 0$)







Training and Evaluation

Models trained using

Trained and evaluated on 2x Nvidia GeForce® Gtx 980 GPUs

Each model was trained fully supervised for 250 epochs

Network Parameter Optimization

- Stratified Mini-batch gradient descent (batch size of 64 examples)
- minimize cross-entropy loss (measure of difference between probability distributions)
- Adaptive Moment Estimation (Adam) optimizer
- Learning Rate Scheduler → decreased learning rate by 10^{-3} after 10 epochs of no improvement
- L^2 regularization with coefficient of 10^{-2}
- 80/20 Train/validation split

Network evaluated using Area Under the Receiver Operating Characteristic Curve

Training and Evaluation

Models trained using

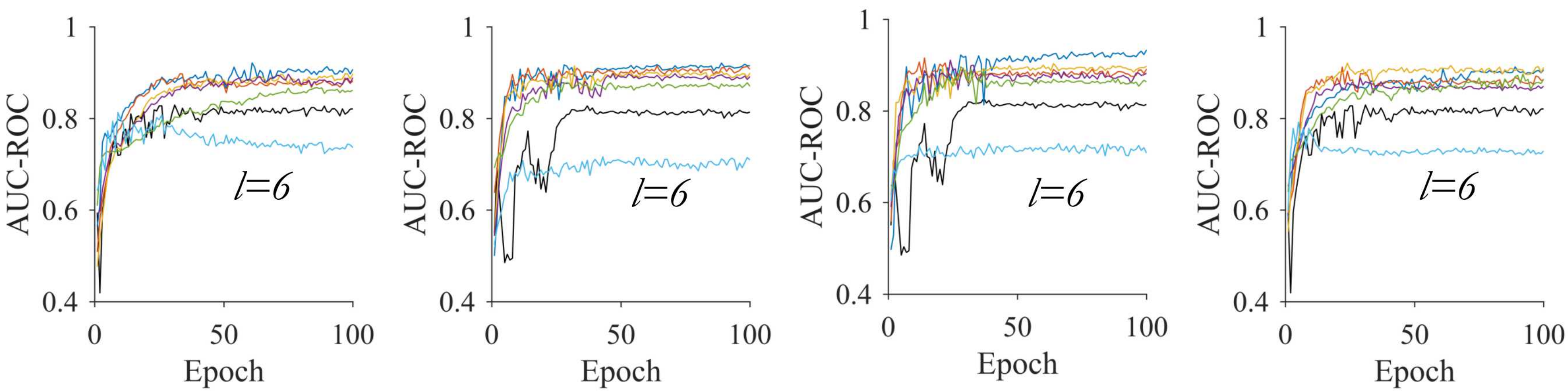
Trained and evaluated on 2x Nvidia GeForce® Gtx 980 GPUs

Each model was trained fully supervised for 250 epochs

Network Parameter Optimization

- Stratified Mini-batch gradient descent (batch size of 64 examples)
- minimize cross-entropy loss (measure of difference between probability distributions)
- Adaptive Moment Estimation (Adam) optimizer
- Learning Rate Scheduler → decreased learning rate by 10^{-3} after 10 epochs of no improvement
- L^2 regularization with coefficient of 10^{-2}
- 80/20 Train/validation split

Network evaluated using Area Under the Receiver Operating Characteristic Curve



Evaluation

Baseline AUC-ROC: 82.9 (Accel) and 82.7 (Accel + Gyro)

Model	**Layers Transferred, *l*, to FallsNet**					
	1	2	3	4	5	6
ParNet(All, Accel)	92.1%	89.7%	90.4%	89.7%	87.0%	81.3%
ParNet(All, Accel + Gyro)	92.1%	91.7%	91.5%	90.1%	87.9%	71.9%
ParNet(Waist, Accel)	91.3%	91.2%	**92.1%**	88.8%	**89.7%**	**79.2%**
ParNet(Waist, Accel + Gyro)	**93.3%**	**91.5%**	90.3%	**91.1%**	86.9%	73.5%

Best AUC results obtained with $l = 1$ and Waist, Accel + Gyro

10.4% improvement in AUC-ROC over baseline

Discussion

Human Activity Recognition

- Inclusion of accelerometer and gyroscope gives better results over accelerometer only
- Training with only acceleration ParNet only learns features related to gait mechanics
- Training with acceleration and gyroscope ParNet learns features related to joint rotation
- Training on data with single placement location generalized better model trained on all placement locations

Transfer Learning and Falls Risk Classification

- ParNet(Waist, Accel + Gyro)with $l = 1$ in terms of AUC-ROC has better transfer learning ability
- Layers 1 and 2 of ParNet appear to sufficiently learn generalized features related to gait
- Performance degrades as deeper layers in ParNet are transferred to FallsNet

Summary

Proposed a deep neural network method for classifying older adults as low/high risk of falling

Trained deep neural network (ParNet) for pedestrian activity recognition

Modified ParNet for falls risk classification (FallsNet)

Applied transfer learning to ParNet

Found that in terms of AUC-ROC only the first layer of ParNet needs to be transferred to FallsNet

Networks trained on waist only data did better than data with all placement locations

Conclusion

Future Work

Data Collection

- Expand data collection to a longitudinal study
  - Include follow-up into regularly defined intervals
  - Use study for validating Bayesian classifier and FallsNet
  - Include data collection (walkway and inertial) with follow up

Falls Risk Classifier to Gait Parameter Estimator

- Train FallsNet to estimate gait parameters from inertial measurements
- Using Bayesian classifier to provide falls risk classification decision
- Inertial Data → Gait Parameter Estimation → Bayesian Classification → Falls Risk Decision

Falls Risk Classification to Falls Risk Anomaly Detection

- Train only on inertial data from individual’s identified as low risk
- Seq2Seq architectures used to like an autoencoder,
  - seq → compressed version → reconstructed seq
- Reconstruction error for each sensor channel as features in classifier

Conclusions

Summary of Dissertation Results

- Demonstrated we can cluster vectors of risk ratio as low/high falls risk using *k*-means
- Showed that a Bayesian classifier we can classify vectors of gait variables as low/high risk while quantifying classifier decision uncertainty
- Show how to pre-train a deep neural network to learn feature representation related to human motion using publicly available pedestrian activity data
- Showed how to use a pre-trained deep neural network as feature extractor for falls risk classification
- Showed how to classify falls risk using inertial gait measurements collected from a smartphone
- End-to-end training of a deep neural network for falls risk classification from inertial measurements of gait

Contributions

- Shown how to use gait variables associated with an increase risk of falling to provide a low/high risk label
- Shown how to "learn" features related to human motion using a deep neural network
- Shown how to apply transfer learning to adapt a pre-trained network for falls risk classification

Backup

Pedestrian Activity Recognition: Evaluation

ParNet(All, Accel)

Actual Class \ Predicted Class	Stay	Walk	Jog	Skip	Up	Down
Stay	527	0	0	0	1	1
Walk	8	498	0	2	11	2
Jog	6	0	530	4	1	3
Skip	5	1	2	538	1	1
Up	6	3	0	0	600	3
Down	9	9	0	1	23	590

ParNet(All, Accel + Gyro)

Actual Class \ Predicted Class	Stay	Walk	Jog	Skip	Up	Down
Stay	528	0	0	0	0	1
Walk	7	503	0	0	9	2
Jog	5	0	533	4	0	2
Skip	4	1	0	539	4	0
Up	6	6	0	0	598	2
Down	7	7	1	4	16	597

ParNet(Waist, Accel)

Actual Class \ Predicted Class	Stay	Walk	Jog	Skip	Up	Down
Stay	347	0	0	0	1	0
Walk	0	345	5	0	2	1
Jog	0	0	309	3	0	1
Skip	5	0	1	330	1	1
Up	5	1	0	0	372	3
Down	1	0	1	1	2	369

ParNet(Waist, Accel + Gyro)

Actual Class \ Predicted Class	Stay	Walk	Jog	Skip	Up	Down
Stay	346	0	1	0	1	0
Walk	0	347	5	0	1	0
Jog	0	0	312	1	0	0
Skip	5	0	0	333	0	0
Up	4	0	0	0	376	1
Down	2	0	1	1	2	368

Evaluation

Baseline Accuracy: 75.8 (Accel) and 75.8 (Accel + Gyro)

Model	Layers Transferred, *l*, to FallsNet					
	1	2	3	4	5	6
ParNet(All, Accel)	80.3%	82.6%	81.1%	75.8%	78.8%	37.1%
ParNet(All, Accel + Gyro)	84.1%	81.8%	78.8%	83.3%	78.8%	68.9%
ParNet(Waist, Accel)	84.1%	**85.6%**	**85.6%**	80.3%	**83.3%**	39.4%
ParNet(Waist, Accel + Gyro)	**84.9%**	76.5%	80.3%	**84.1%**	81.1%	64.4%

Best accuracy results obtained with *l* = 2 and Waist, Accel

9.8% improvement in accuracy over baseline

Evaluation

Baseline Specificity (1 - FPR): 76.0 (Accel) and 78.0 (Accel + Gyro)
ROC threshold set such that TPR == (1 – FPR)

Model	**Layers Transferred, *l*, to FallsNet**					
	1	2	3	4	5	6
ParNet(All, Accel)	81.0%	77.0%	79.0%	80.0%	77.0%	76.0%
ParNet(All, Accel + Gyro)	83.0%	80.0%	81.0%	83.0%	78.0%	68.0%
ParNet(Waist, Accel)	82.9%	**85.0%**	**85.0%**	80.0%	**80.0%**	**73.0%**
ParNet(Waist, Accel + Gyro)	**84.0%**	81.0%	80.0%	**84.0%**	80.0%	65.0%

Best specificity results obtained with *l* = 2, 3 and Waist, Accel

9% improvement in specificity over baseline