* Gradient-based methods are subject to gradient shattering issue.
* Sensitivity analysis
  + What is the gradient for every single pixel
  + Explains a variation of the function not the function itself.
  + Input gradient becomes highly varying and unreliable as the neural network depth increases.
  + Where as in gradient shattering (sensitivity/saliency), the model pays attention to the road, because the question is that how I can wiggle pixels such that I get a scooter. “Which pixels lead to increase/decrease of prediction score when changed”
  + What do I have to do to this pixel in order to make this a scooter. There’ll be a lot of noise because of gradient shattering.
* Heatmap is the output of LRP algorithm:
  + Blue speaks against sheep, why it is not 9, least relevant work
  + Red speaks for sheep, why it is 3, most relevant word
* LRP works well for deep models, recurrent deep models (LSTM), kernel methods, ...
  + You can convert a kernel to a neural network or a model.
* Heatmapping: LRP is not subject to gradient shattering.
  + It can analyze any network that is available, e.g AlexNet
  + Scooters
    - Taillight, tire, …
  + LRP tries to decompose the function given by neural network.
    - Which pixels contribute how much to the classification.
    - Deep Taylor Decomposition
* Applications
  + Comparing models
    - Text classification
      * Word2Vec: CNN
        + Semantic meaning
        + Can score the words that have not been in the training dataset
      * BoW: SVM
        + Word statistics
        + Assigns relevance to the words not relevant to the target class
    - Computer vision
      * BVLC Reference CaffeNet
      * GoogleNet
        + Better performance
        + More layers
        + Sparse
        + Focuses on the face
        + Does not focus on the background
  + Quantify context use (how much the model uses the context)
    - Importance of context = relevance outside bbox/ relevance inside bbox
    - For some objects, context is more important
      * Indoor objects (sofa)
    - For some, context is not that important (airplane)
    - Assigning more importance to context leads to poorer performance
  + Compare configurations
    - Face analysis
      * Gender classification
        + With pre-training

Chin and beard when classifying as man

Eye and hair when classifying as woman

* + - * + Without pre-training the heatmaps are not intuitive.
      * Age classification
        + Presence and absence of laughing for classification
      * Face morphing attack
  + Learn new representations
    - Uses PCA.
    - Compute the weighted average of the relevance scores of the words to represent the whole document in a new way.
  + Understand the model and gain new insight
    - Image classification
      * Classify as boat
        + Focus is on the water
      * Classify as train
        + Focus is on the track
    - Motion detection
      * Chewing
        + Movements of the whole head
    - Sentiment classification
      * LSTM model
      * Understands negation
        + Red and blue
    - Video classification
      * Explanation of frames
      * Relevance of each frame
      * Focuses on the beginning and the end of the frame
        + Outer parts of a video
        + No uniform distributed relevance
      * Tried the video by fast-forwarding to see more of the video
        + Accuracy increased.
    - Speech analysis
      * Uses spectrograms
      * Harmonics (spacing between frequencies) are an important factor to classify as male or female speaker.
    - VQA (Visual Question Answering)
      * Text:
        + Words like “behind”
        + Count:

“how” or “how many”

* + - * + Size:

Size related words

* + - * Image:
        + Object under question
    - Atari game
      * LRP:
        + Where the ball is
      * Sensitivity
        + When the ball could be in the next frame
      * You can evaluate the model epoch-wise after n iterations
      * Pinball
        + Ball
        + Paddle
        + Tunnel
* Comparing classifiers (finding flaws in models)
  + Same performance error for horse class in both models (FV and DNN). Why?
    - FV uses (artefacts) the image tag to classify (cheating)
* Neuroscience
  + Brain computer interfacing
    - Measuring data from the brain and decode the brain states in real-time to control things: computer games, wheelchairs, …
    - What actually happens in the decoder?
* Chemistry, physics, materials
* Learning Atomistic Representations with Deep Tensor Neural Networks
* Machine Learning for Morpho-Molecular Integration
* Support Vector Data Description
  + Explaining one-class
  + Can detect anomalies in images that edge detector cannot