Multimodal Machine Learning

Michael Eisele 7089760

Universität Hamburg

DER FORSCHUNG | DER LEHRE | DER BILDUNG

Definition

Modality

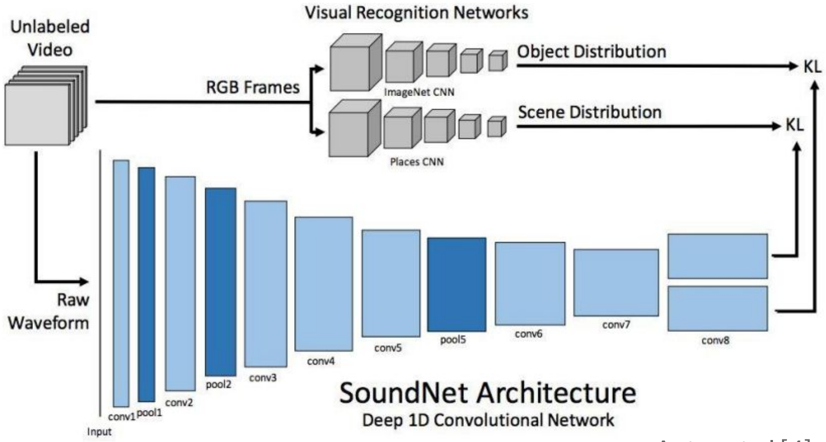
A particular way of doing or experiencing something. It refers to a certain type of information and/or the representation format in which information is stored. [1]

Motivation

- Similar to how humans perceive their environment
- Some information is difficult to acquire unimodally
- Increased robustness

Examples of MMML applications

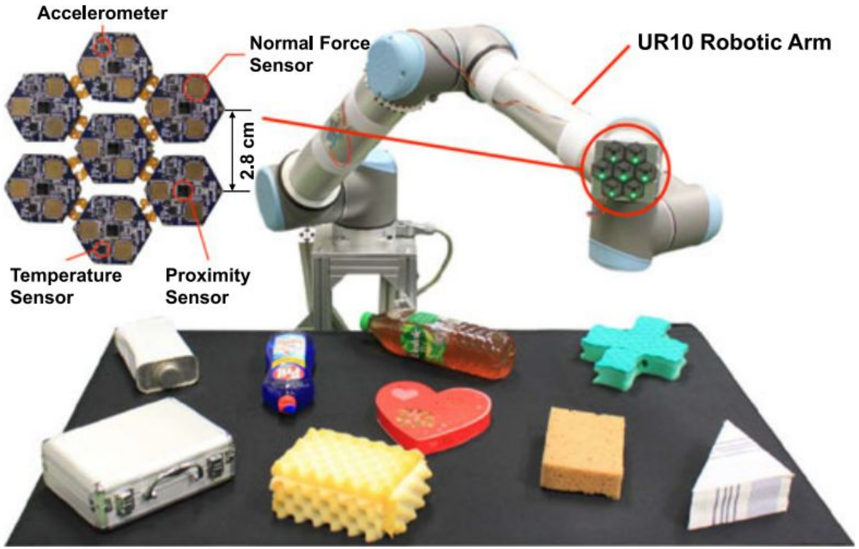
- Natural language processing/ Text-to-speech
- Image tagging or captioning [3]
- SoundNet recognizing objects from sound [4]



Aytar et al [4]

Robot learning to grasp objects, Kaboli et al [2]

- Starts estimating object positions in a grid
- Pressing, sliding, static contact to explore physical properties
- Tries to minimize training for unknown objects



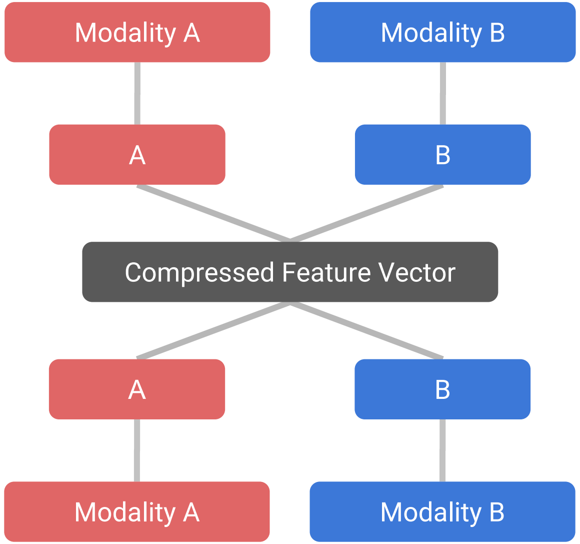
Technical Challenges for MMML [5]

- Representation
- Fusion
- Alignment
- (Translation)
- (Co-Learning)

Representation

Joint representations

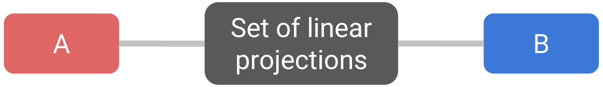
1. Multimodal Tensor Fusion Network
2. **Deep Multimodal autoencoders**
3. Deep Multimodal Boltzmann machines

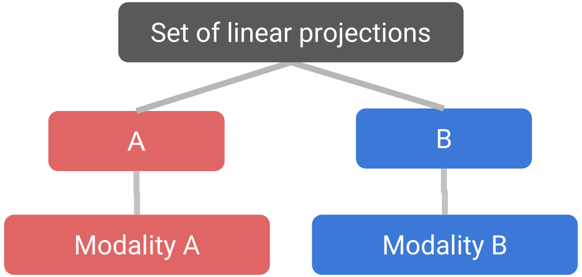


Representation

Coordinated representations

1. Multimodal Vector Space Arithmetic
2. **Deep Canonical Correlation Analysis**
3. Deep Canonically Correlated Autoencoders

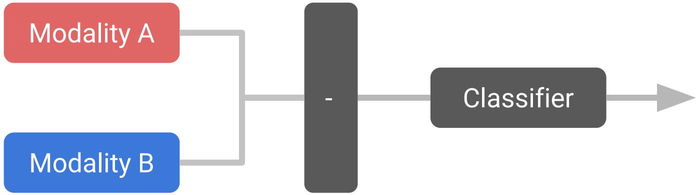




Fusion

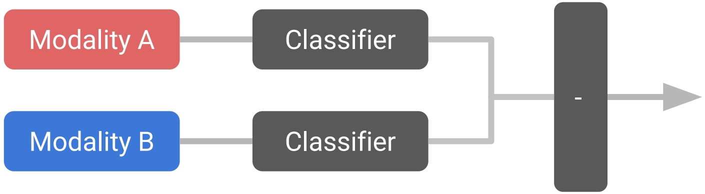
Model-Agnostic Approaches

1. **Early Fusion**
2. **Late Fusion**
3. Hybrid Fusion



Model-Based Approaches

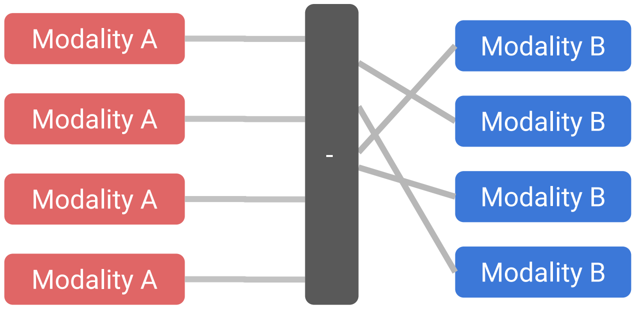
1. Multiple Kernel Learning
2. Deep Neural Network



Alignment

Explicit Alignment
Goal is to find correspondences between elements

Implicit Alignment
Goal is to solve a given problem



The McGurk Effect

The McGurk Effect in MMML

Audio / Visual Setting	Model prediction		
	/ga/	**/ba/**	**/da/**
Visual **/ga/**, Audio **/ga/**	82.6%	2.2%	15.2%
Visual **/ba/**, Audio **/ba/**	4.4%	89.1%	6.5%
Visual **/ga/**, Audio **/ba/**	28.3%	13.0%	58.7%

Sources

Multimodal Machine Learning: A Survey and Taxonomy
https://arxiv.org/pdf/1705.09406.pdf

Multimodal Deep Learning for Robust RGB-D Object Recognition
https://arxiv.org/pdf/1507.06821.pdf

Multimodal Integration Learning of Object Manipulation Behaviors using Deep Neural Networks
https://pdfs.semanticscholar.org/d477/5ba945e93d9ac09fc606ef566fca7c8524e4.pdf

Robot gains Social Intelligence through Multimodal Deep Reinforcement Learning
https://arxiv.org/pdf/1702.07492.pdf

Tutorial on Multimodal Machine Learning
https://www.microsoft.com/en-us/research/wp-content/uploads/2017/07/Integrative_AI_Louis_Philippe_Morency.pdf

Sources

Cambridge Dictionary
https://dictionary.cambridge.org/de/worterbuch/englisch/modality

A Tactile-Based Framework for Active Object Learning and Discrimination using Multimodal Robotic Skin
https://arxiv.org/pdf/1301.3592.pdf

Deep Learning for Detecting Robotic Grasps
https://ieeexplore.ieee.org/stamp/stamp.jsp?arnumber=7961193

Tensor Fusion Network for Multimodal Sentiment Analysis
https://www.aclweb.org/anthology/D17-1115

Feedback & Questions