Eidgenössische Technische Hochschule Zürich
Swiss Federal Institute of Technology Zurich

ICU Cockpit
artificial intelligence for neurointensive care

Not to Cry Wolf:
Distantly Supervised Multitask Learning in Critical Care

Patrick Schwab[1]

@schwabpa

Emanuela Keller[2], Carl Muroi[2], David J. Mack[2], Christian Strässle[2] and Walter Karlen[1]

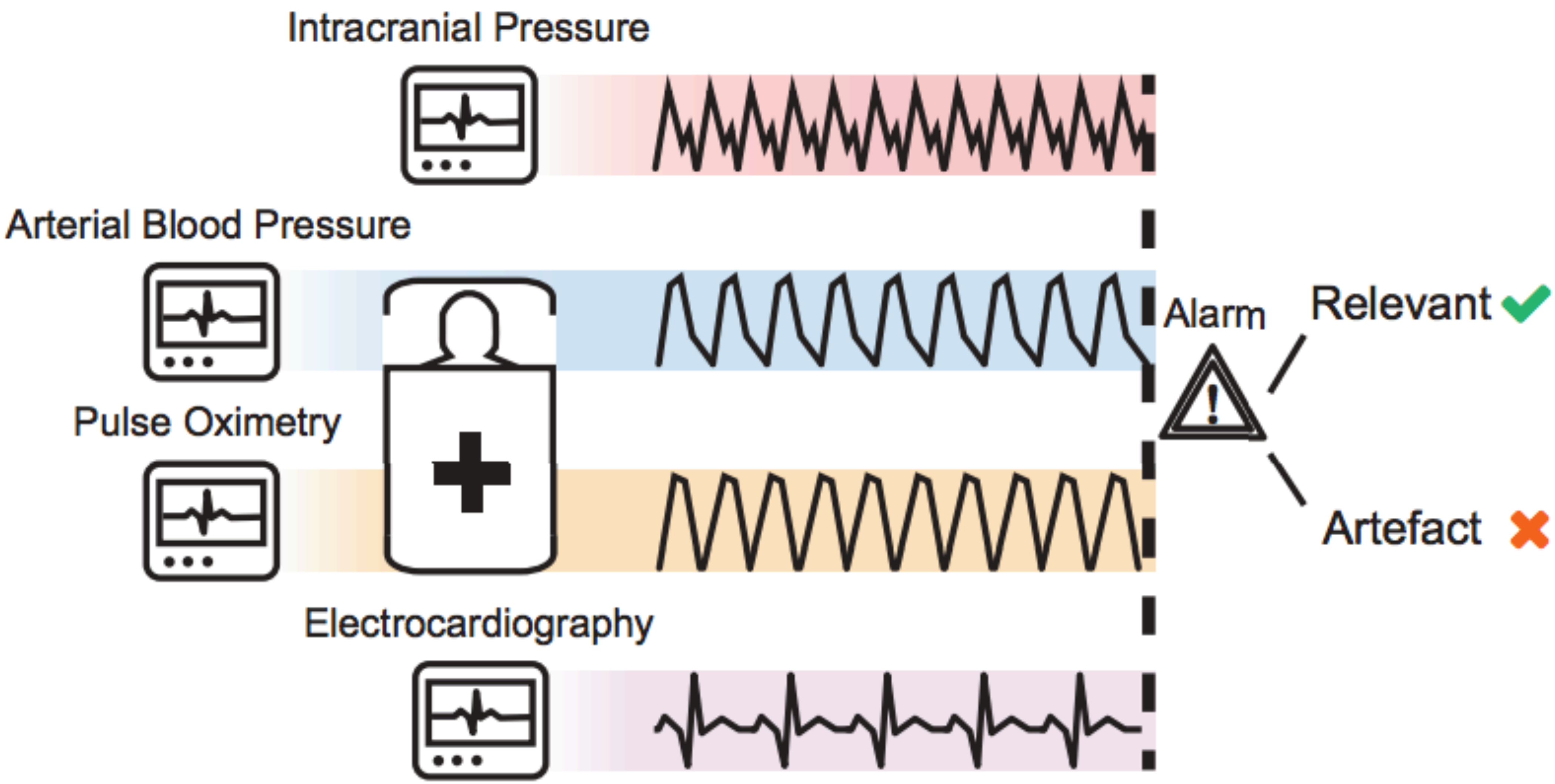
[1]Institute of Robotics and Intelligent Systems, ETH Zurich
[2] Neurocritical Care Unit, University Hospital Zurich

75 NRP Big Data National Research Programme

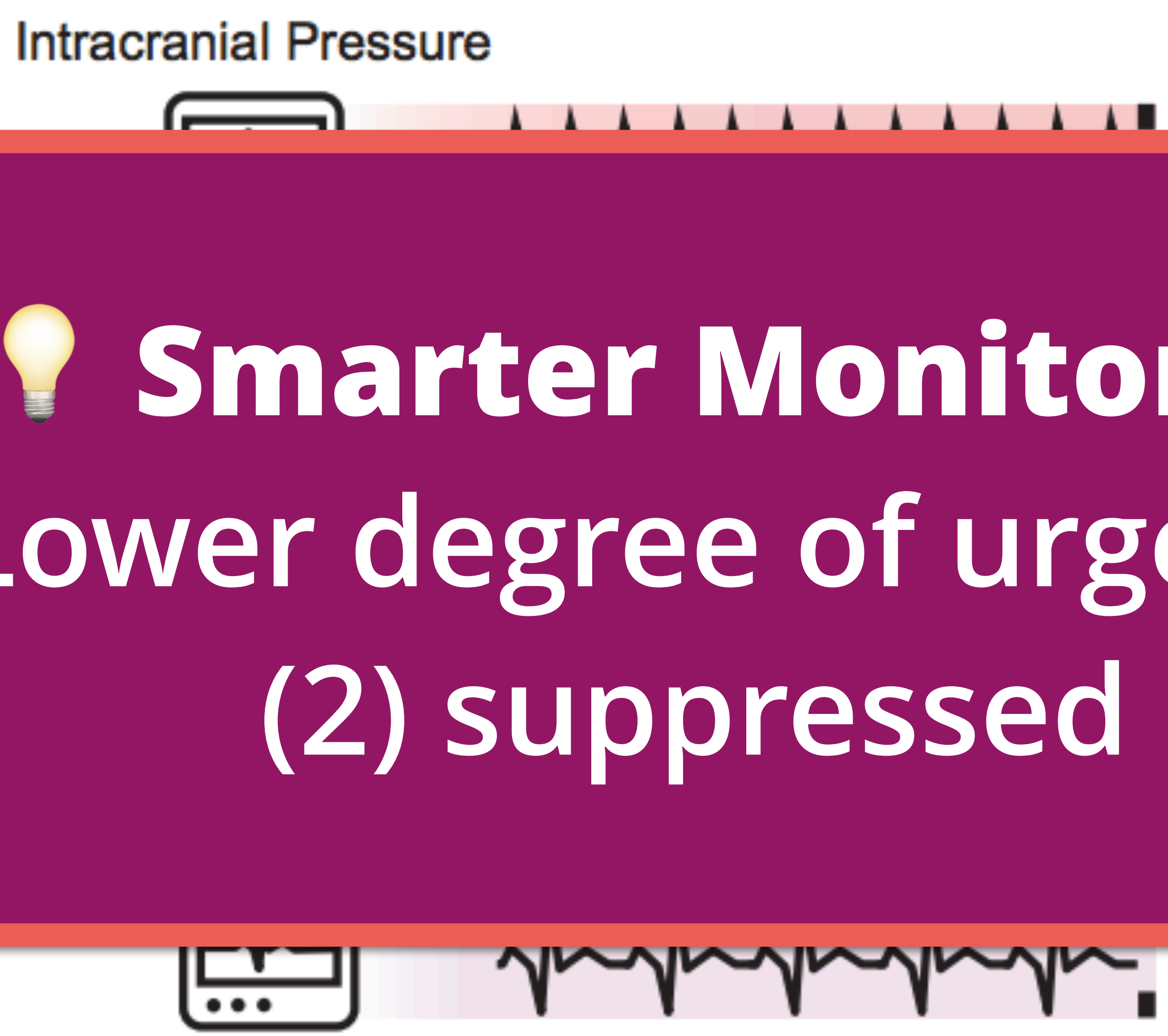
University of Zurich UZH

How Can We Help?

The Idea


Intracranial Pressure
Arterial Blood Pressure
Pulse Oximetry
Electrocardiography
Alarm
Relevant
Artefact

The Idea



💡 **Smarter Monitoring**

(1) Lower degree of urgency, or

(2) suppressed

Challenges

- **Large amounts** of biosignal monitoring data and alarms available
  - But only a limited amount of labelled data
  - Expert labels **expensive** and **time-consuming**
- **Can we make due with a smaller number of labels?**

Semi-supervised Learning

Existing Approaches

- Existing methods to semi-supervised learning in deep networks are roughly:

  1. **Distant / self / weak supervision**
     - **e.g. temporal ensembling[1]**
  2. **Reconstruction-based objectives**
     - **e.g. AE, VAE, Ladder Nets**
  3. **Adversarial learning**
     - **e.g. Feature Matching GANs, CatGAN, Triple-GAN, …**

1 Laine & Aila, ICLR 2017

Existing Approaches

- Existing methods to semi-supervised learning in deep networks are roughly:

1. Distant / self / weak supervision

- **e.g. temporal ensembling[1]**

2. Reconstruction-based objectives

- **e.g. AE, VAE, Ladder Nets**

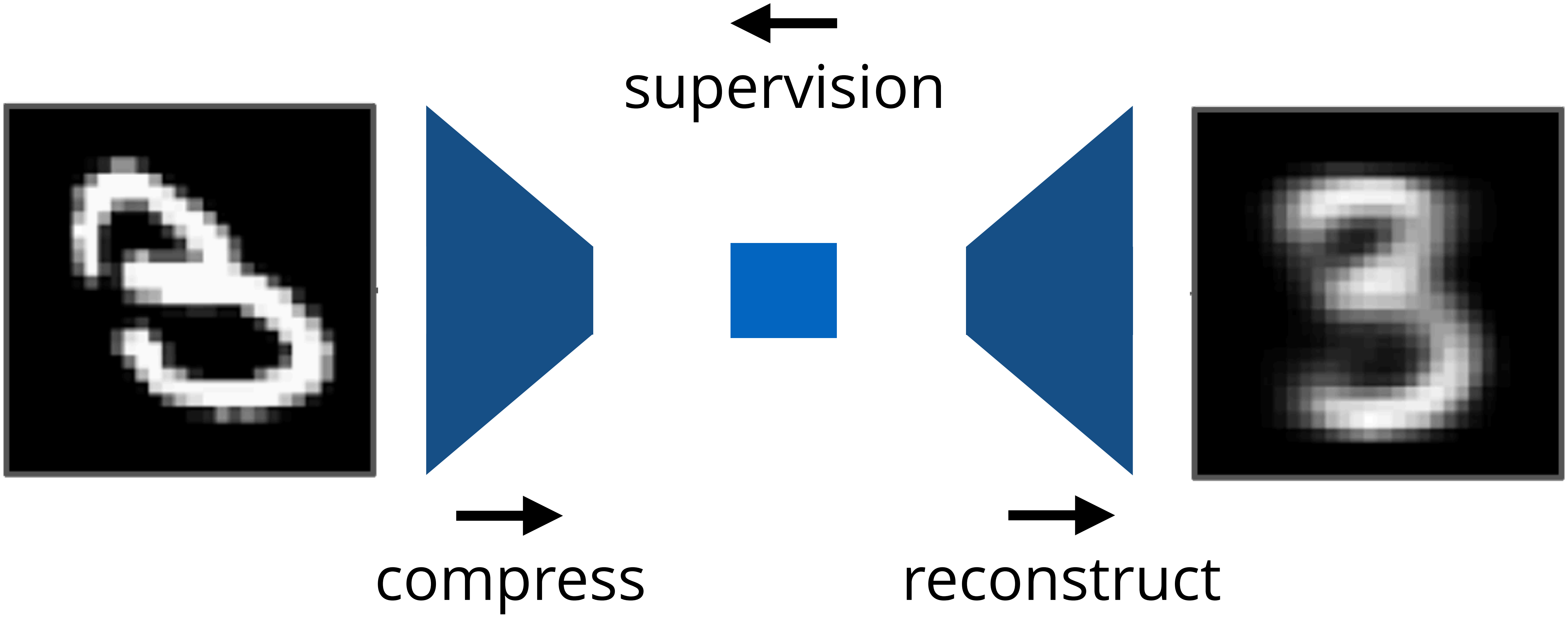
3. Adversarial learning

- **e.g. Feature Matching GANs, CatGAN, Triple-GAN, …**

1 Laine & Aila, ICLR 2017

A Unified View

- **Reconstruction-based SSL** can be viewed as distant supervision where **reconstruction** is the **auxiliary task**



A Unified View

- **Reconstruction-based SSL** can be viewed as distant supervision where **reconstruction** is the **auxiliary task**
- Reconstruction is a **convenient** auxiliary task
    - .. generalises to **all** kinds of models, input data

A Unified View

- **Reconstruction-based SSL** can be viewed as distant supervision where **reconstruction** is the **auxiliary task**
- Reconstruction is a **convenient** auxiliary task
    - .. generalises to **all** kinds of models, input data
- But is it the **best**?

Hypotheses

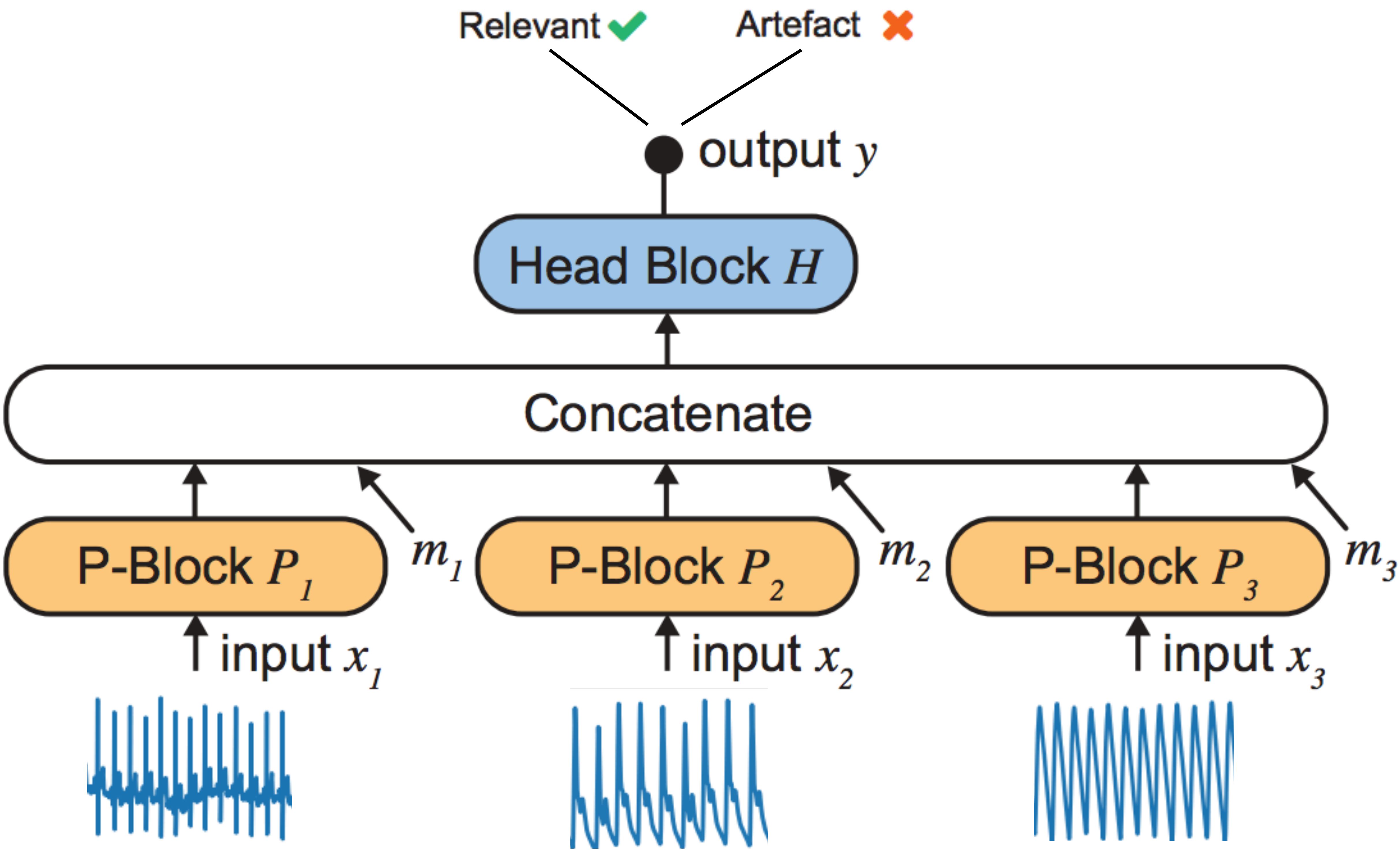
- Recent **empirical successes**[1] with **specifically engineered auxiliary tasks** lead to hypotheses:

(1) More **“related” auxiliary tasks** might be a better choice than reconstruction

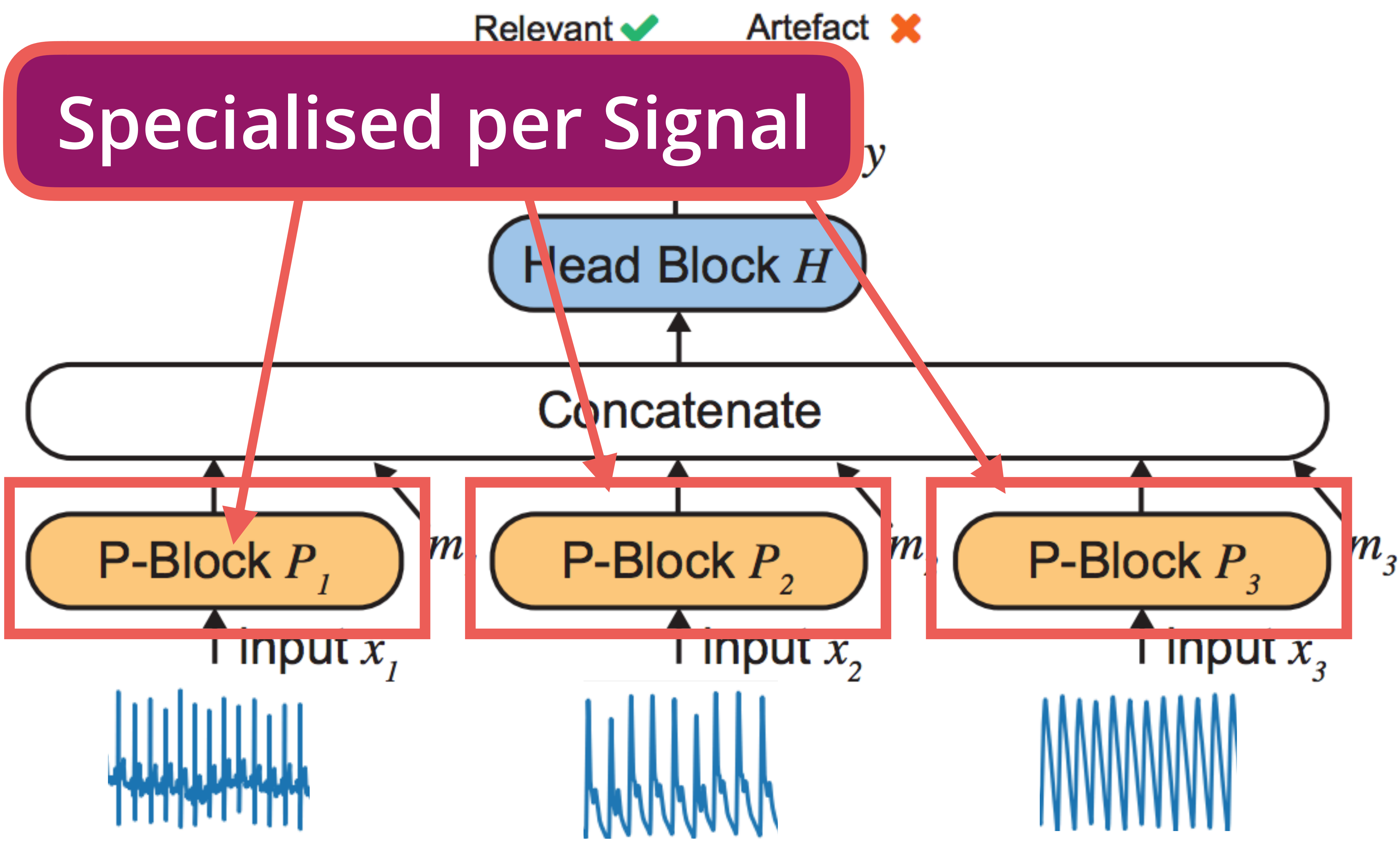
(2) Using **multiple diverse auxiliary tasks** might be better than just one

1 Oquab et al., 2015; Deriu et al., 2017; Doersch & Zisserman, 2017

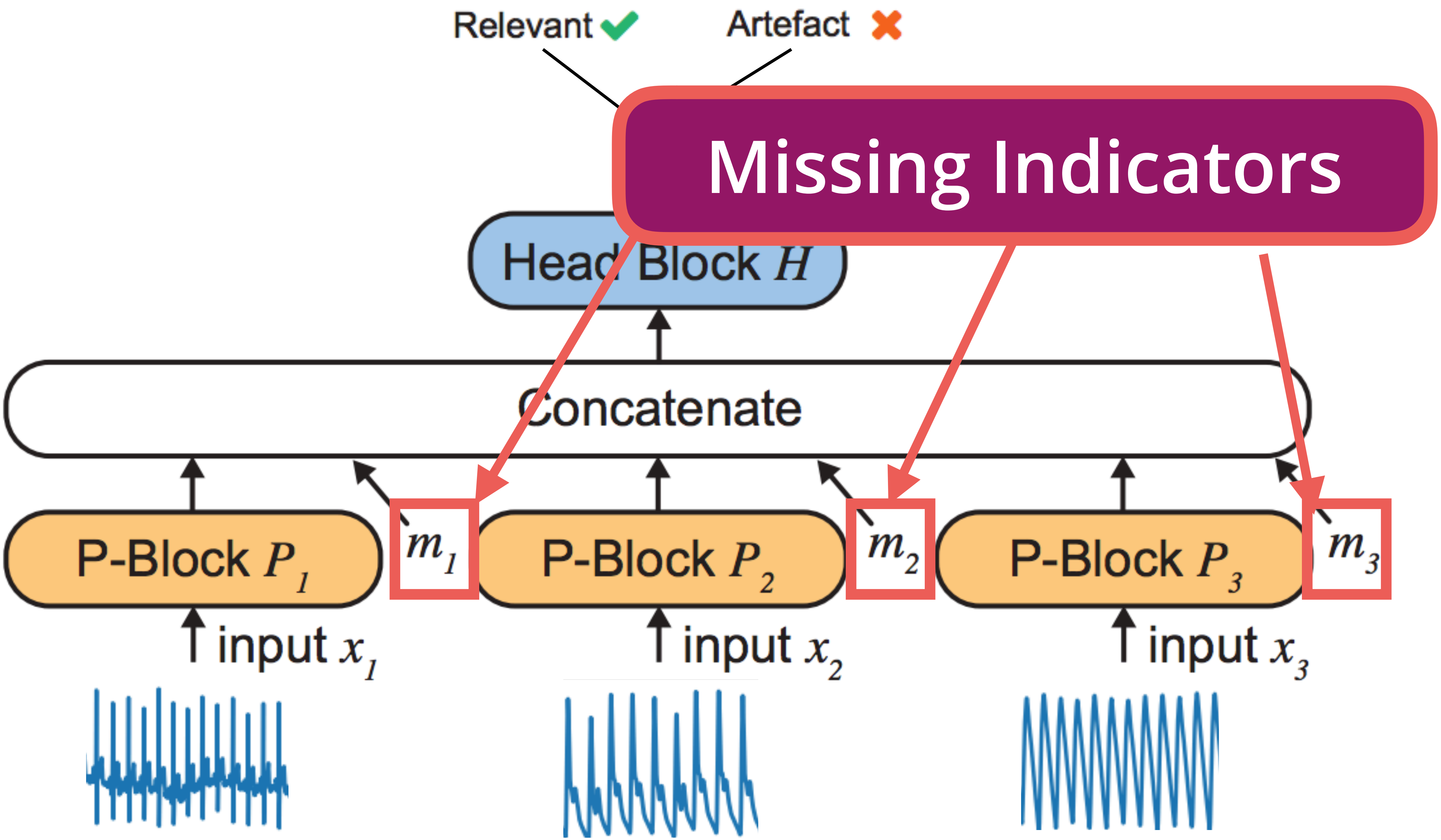
Supervised Learning



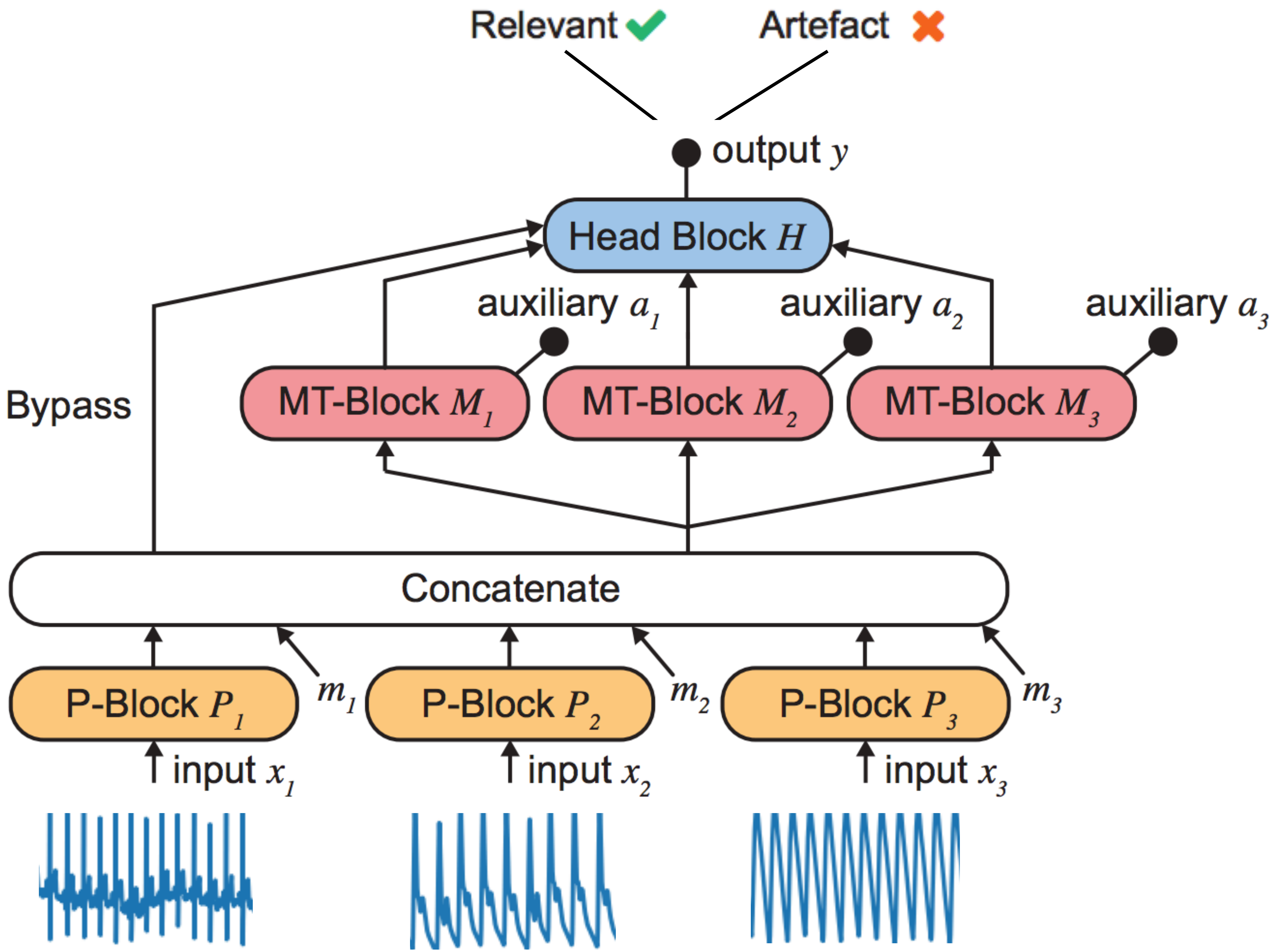
Supervised Learning



Supervised Learning



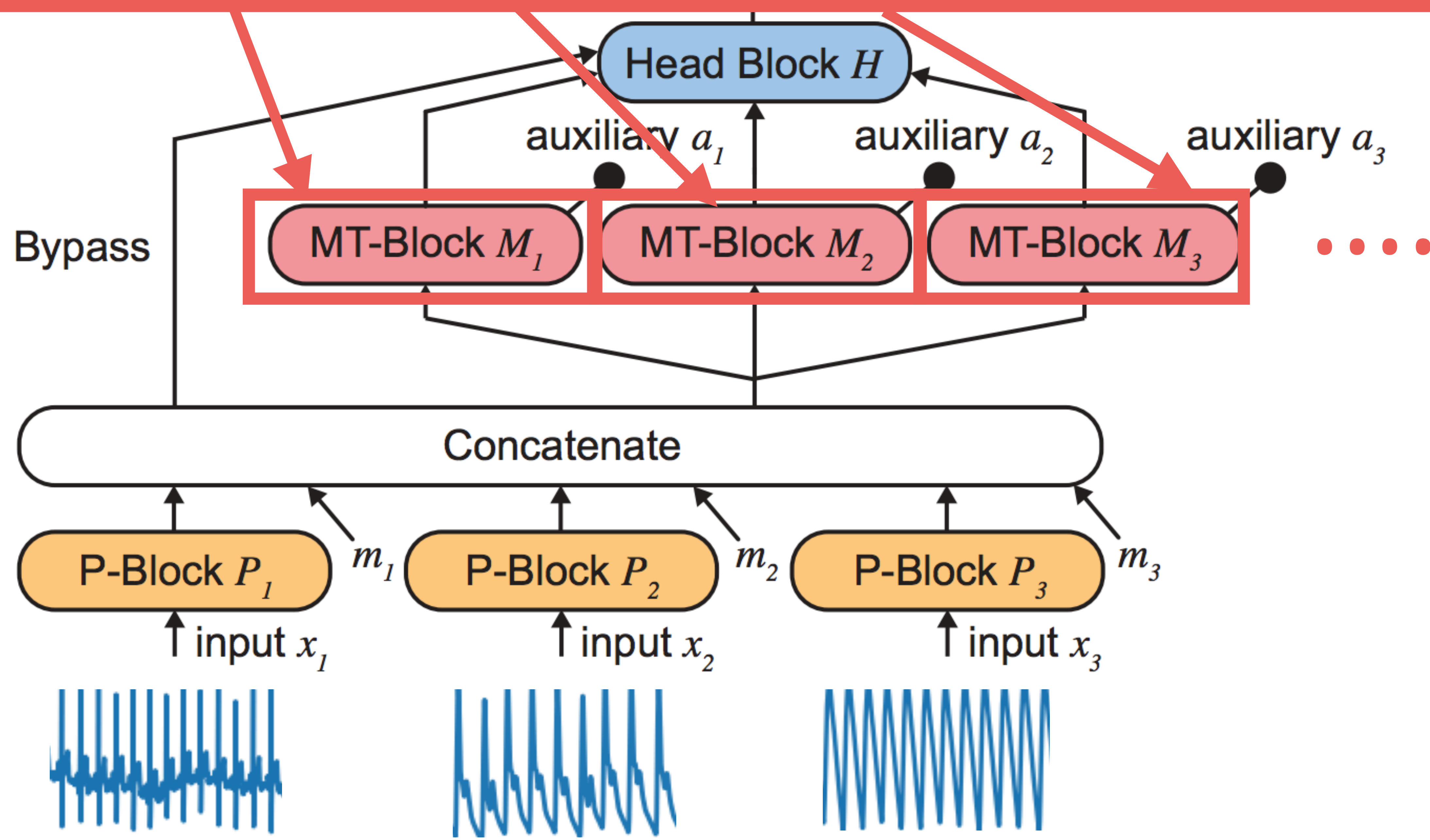
DSMT-Net



Relevant
Artefact
output y
Head Block H
auxiliary a_1
auxiliary a_2
auxiliary a_3
Bypass
MT-Block M_1
MT-Block M_2
MT-Block M_3
Concatenate
m_1
m_2
m_3
P-Block P_1
P-Block P_2
P-Block P_3
input x_1
input x_2
input x_3

DSMT-Net

Any Number of Multitask Blocks



So far so good, but …

1 - **Where could we get a large number of auxiliary tasks from?**

2 - **What about potential adverse interactions between gradients from all these auxiliary tasks?**

1 - Large-scale Auxiliary Task Selection

- **How** do we **select auxiliary tasks** for distant supervision?
  - Identification of **relevant** features in **large feature repository** (auto-corr., power spectral densities..)
    - relevant = **significant correlation**[1] with labels
  - Simple strategies:

    (1) At **random** out of the relevant set, and

    (2) in order of **importance**

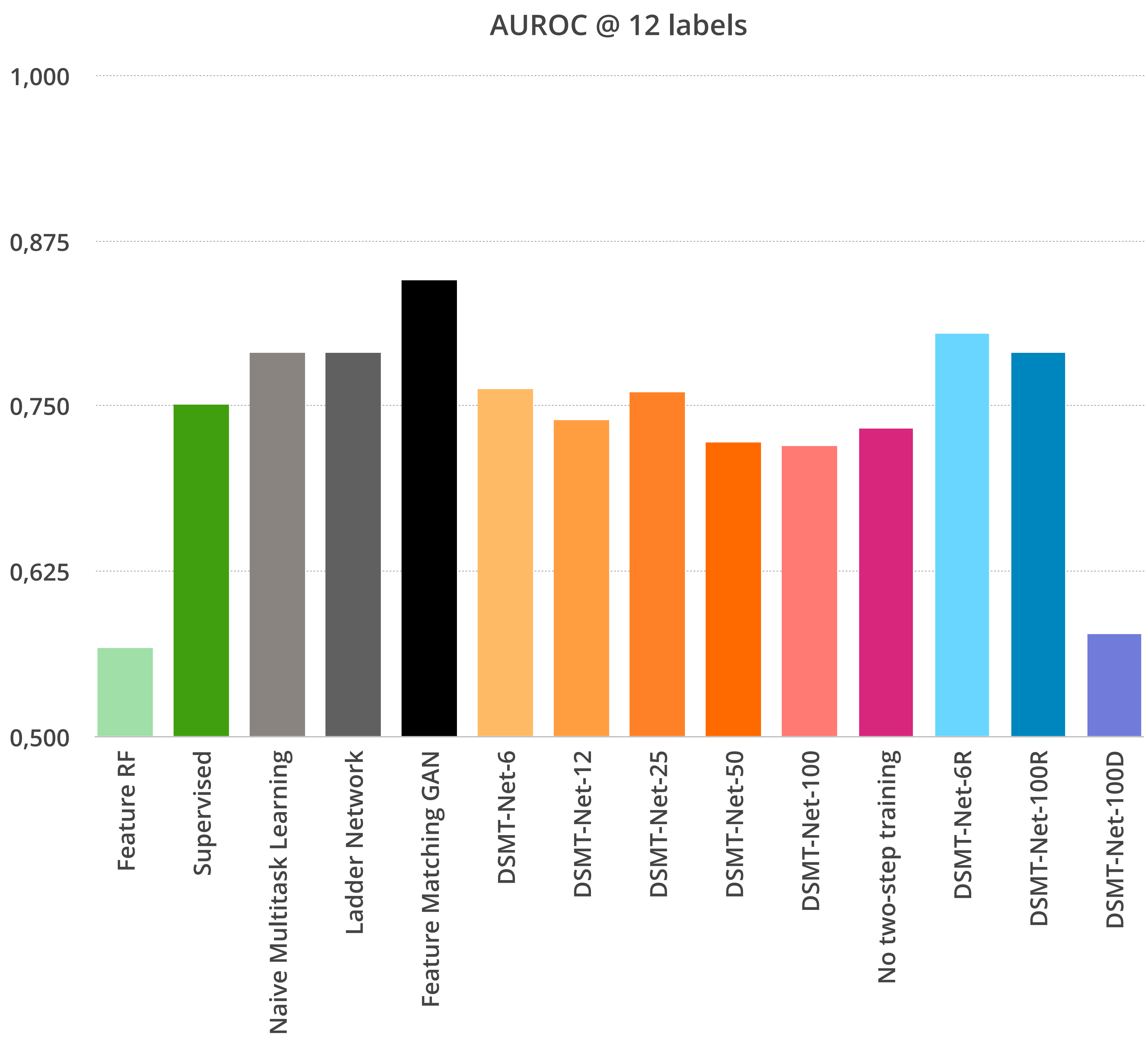
1 Kendall's τ

2 - Combating Adverse Gradient Interactions

- A key issue in end-to-end multitask learning are **adverse gradient interactions**
- We therefore **disentangle** training unsupervised and supervised tasks
- Train in **alternating fashion** in each epoch
  - First unsupervised tasks then supervised tasks
- Similar to alternating training regime in GANs

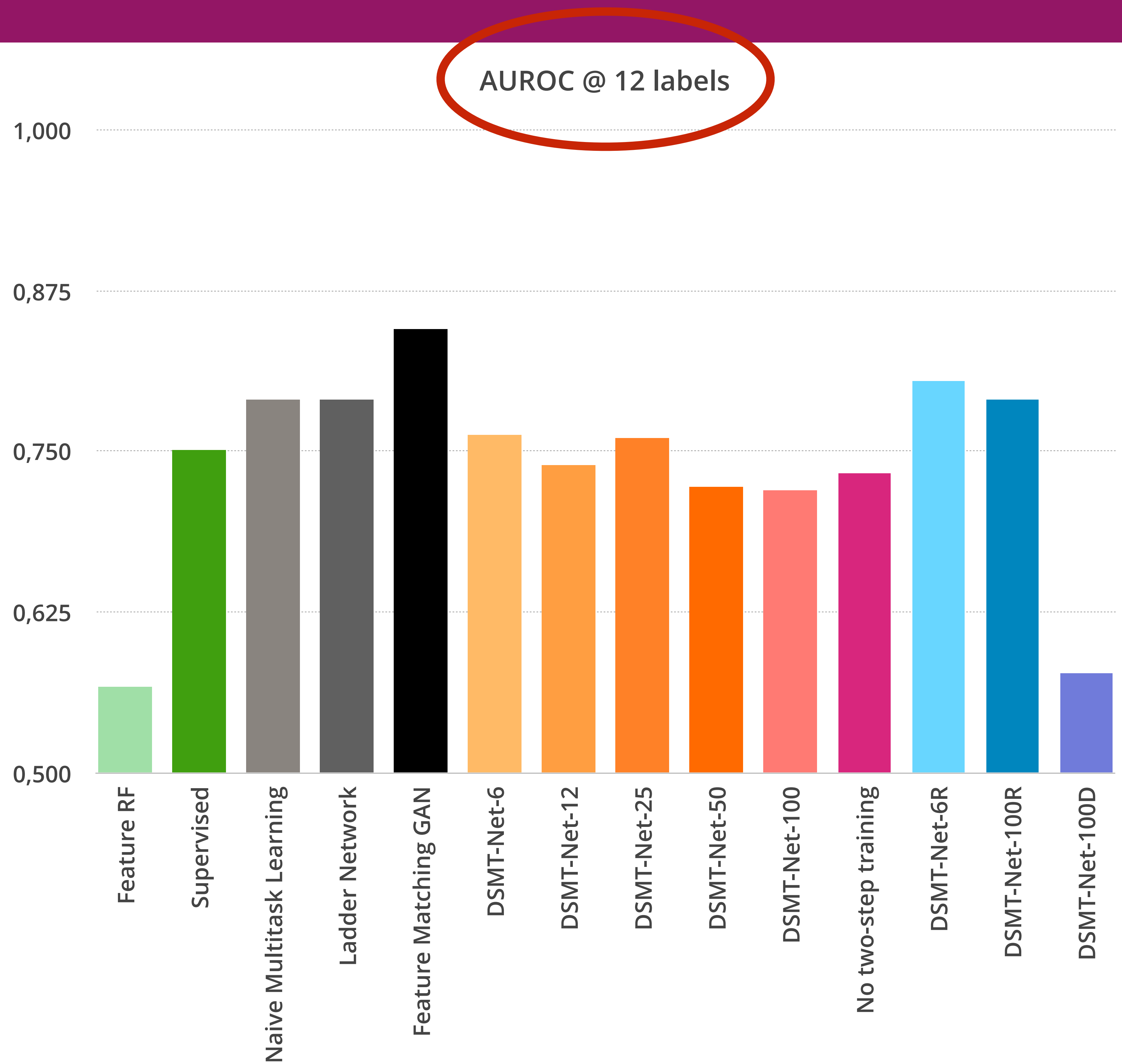
Evaluation

Results



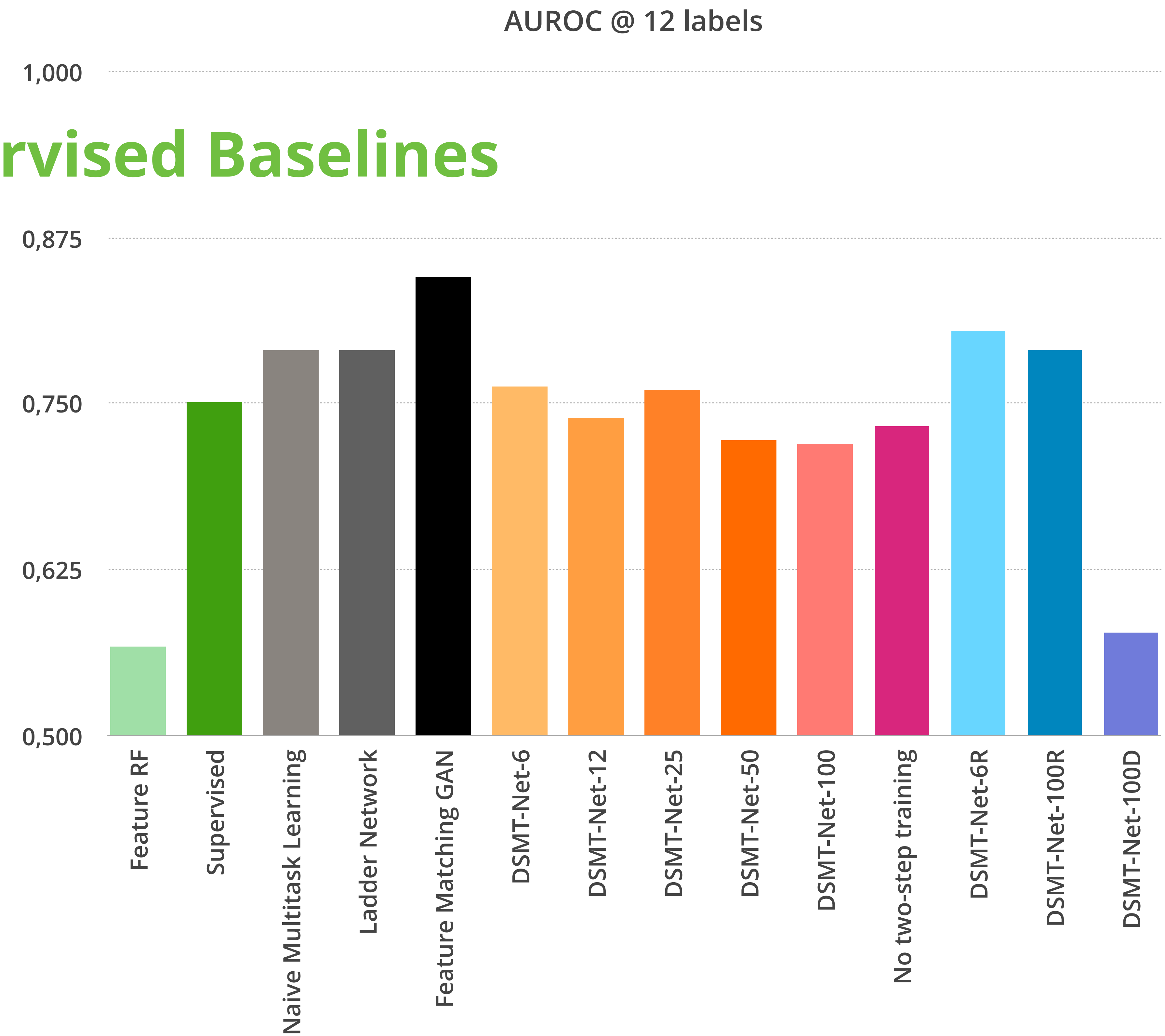
AUROC @ 12 labels
1,000
0,875
0,750
0,625
0,500
Feature RF
Supervised
Naive Multitask Learning
Ladder Network
Feature Matching GAN
DSMT-Net-6
DSMT-Net-12
DSMT-Net-25
DSMT-Net-50
DSMT-Net-100
No two-step training
DSMT-Net-6R
DSMT-Net-100R
DSMT-Net-100D

Results

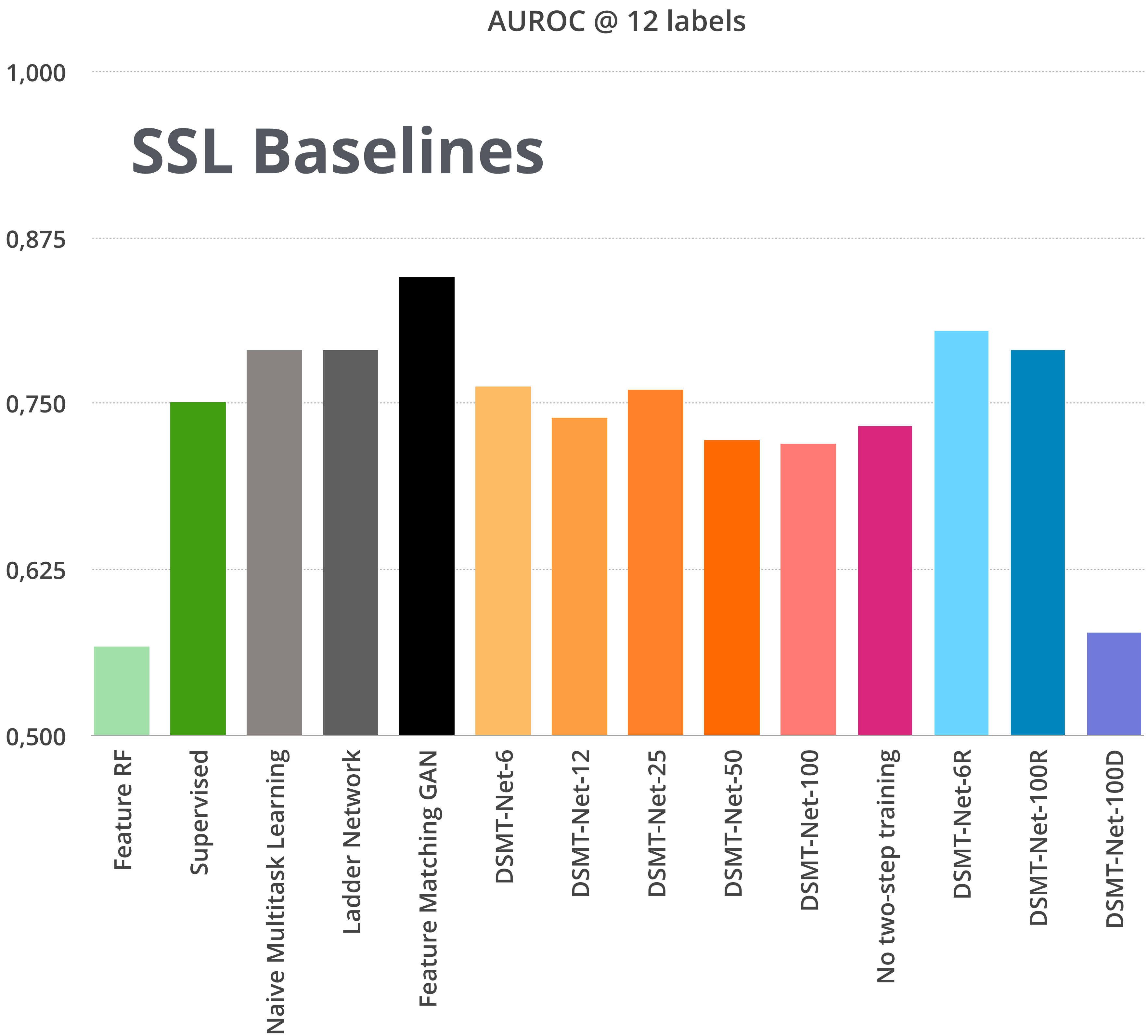

AUROC @ 12 labels
1,000
0,875
0,750
0,625
0,500
Feature RF
Supervised
Naive Multitask Learning
Ladder Network
Feature Matching GAN
DSMT-Net-6
DSMT-Net-12
DSMT-Net-25
DSMT-Net-50
DSMT-Net-100
No two-step training
DSMT-Net-6R
DSMT-Net-100R
DSMT-Net-100D

Results

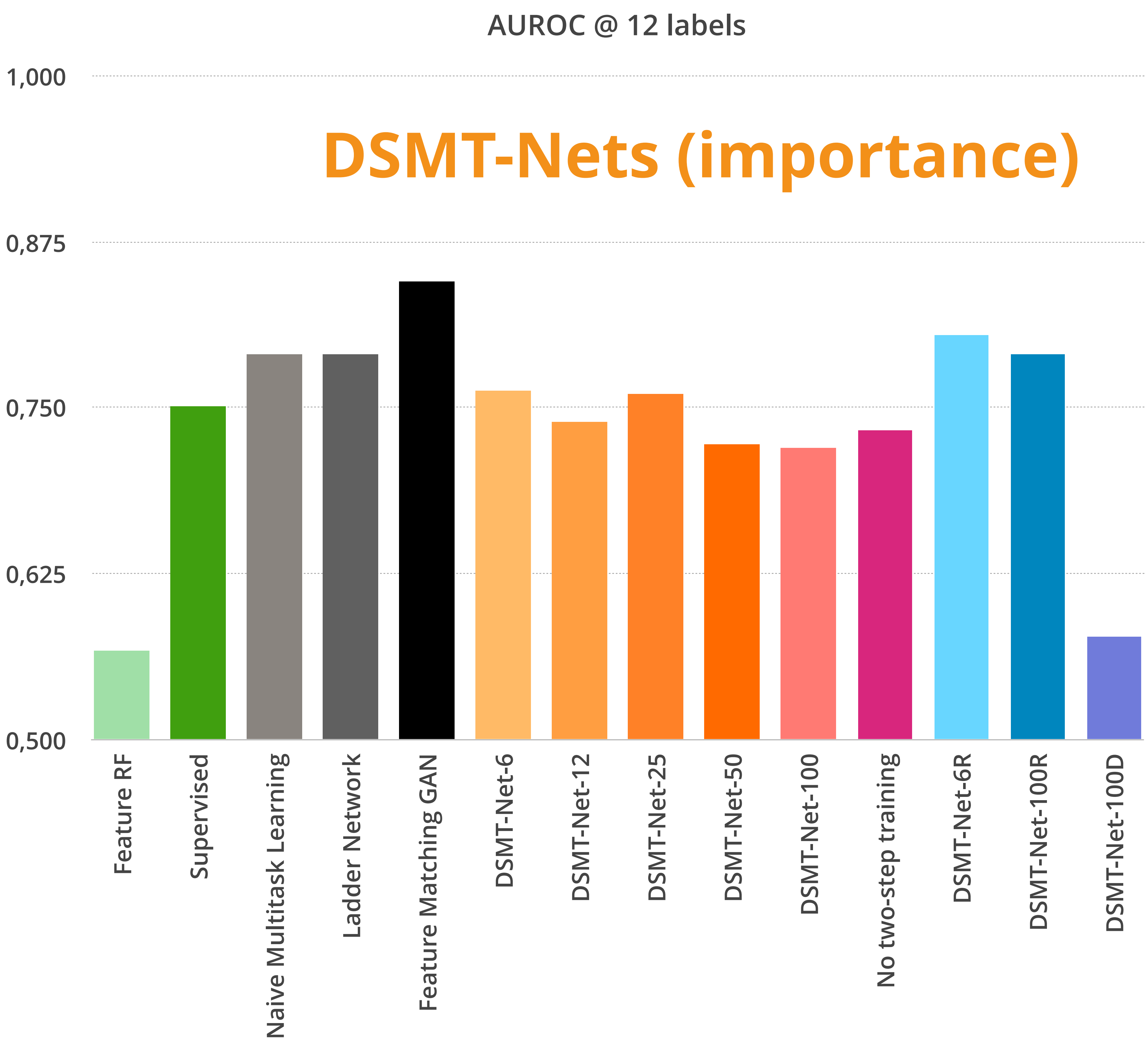
Supervised Baselines



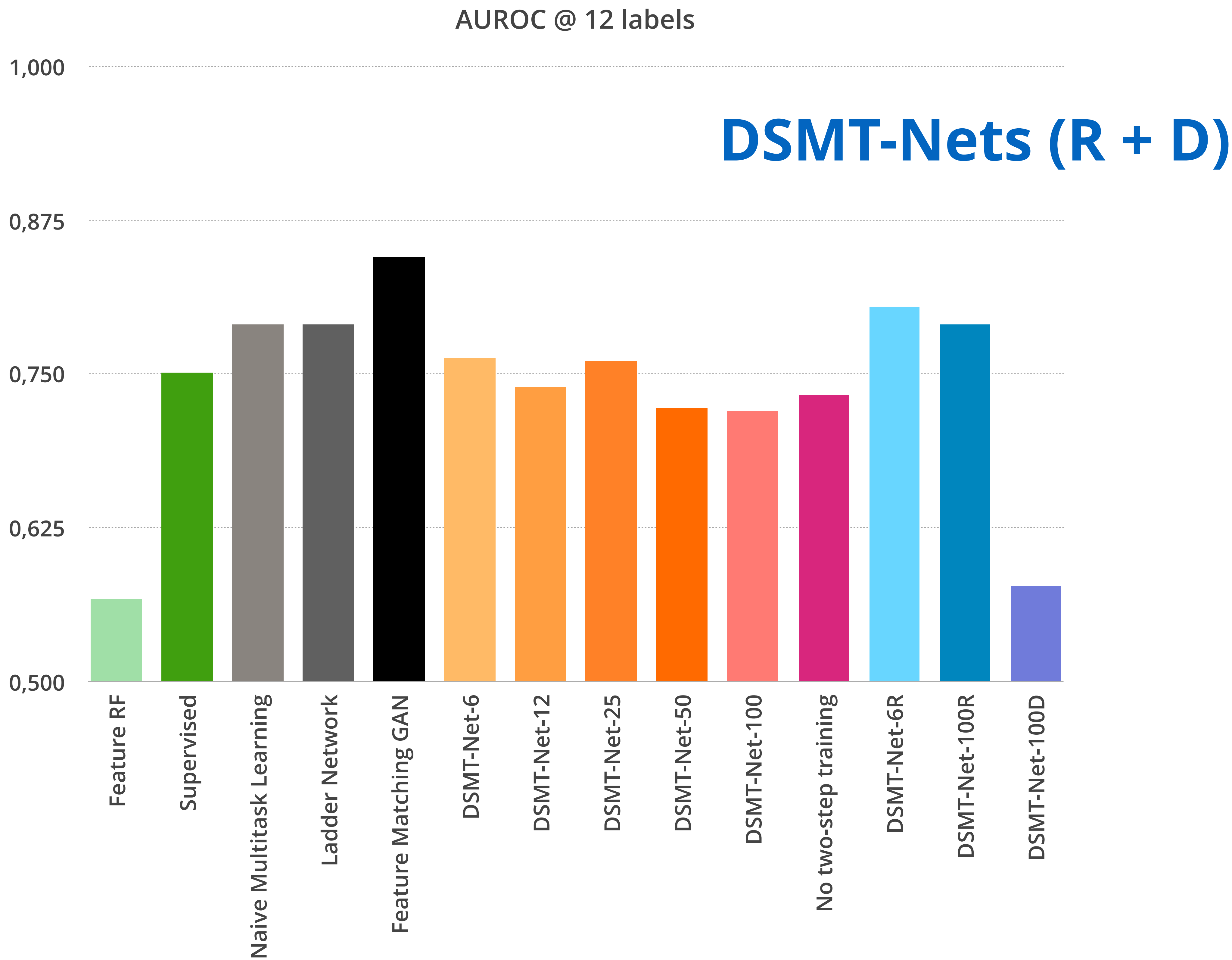
Results


AUROC @ 12 labels
1,000
SSL Baselines
0,875
0,750
0,625
0,500
Feature RF
Supervised
Naive Multitask Learning
Ladder Network
Feature Matching GAN
DSMT-Net-6
DSMT-Net-12
DSMT-Net-25
DSMT-Net-50
DSMT-Net-100
No two-step training
DSMT-Net-6R
DSMT-Net-100R
DSMT-Net-100D

Results

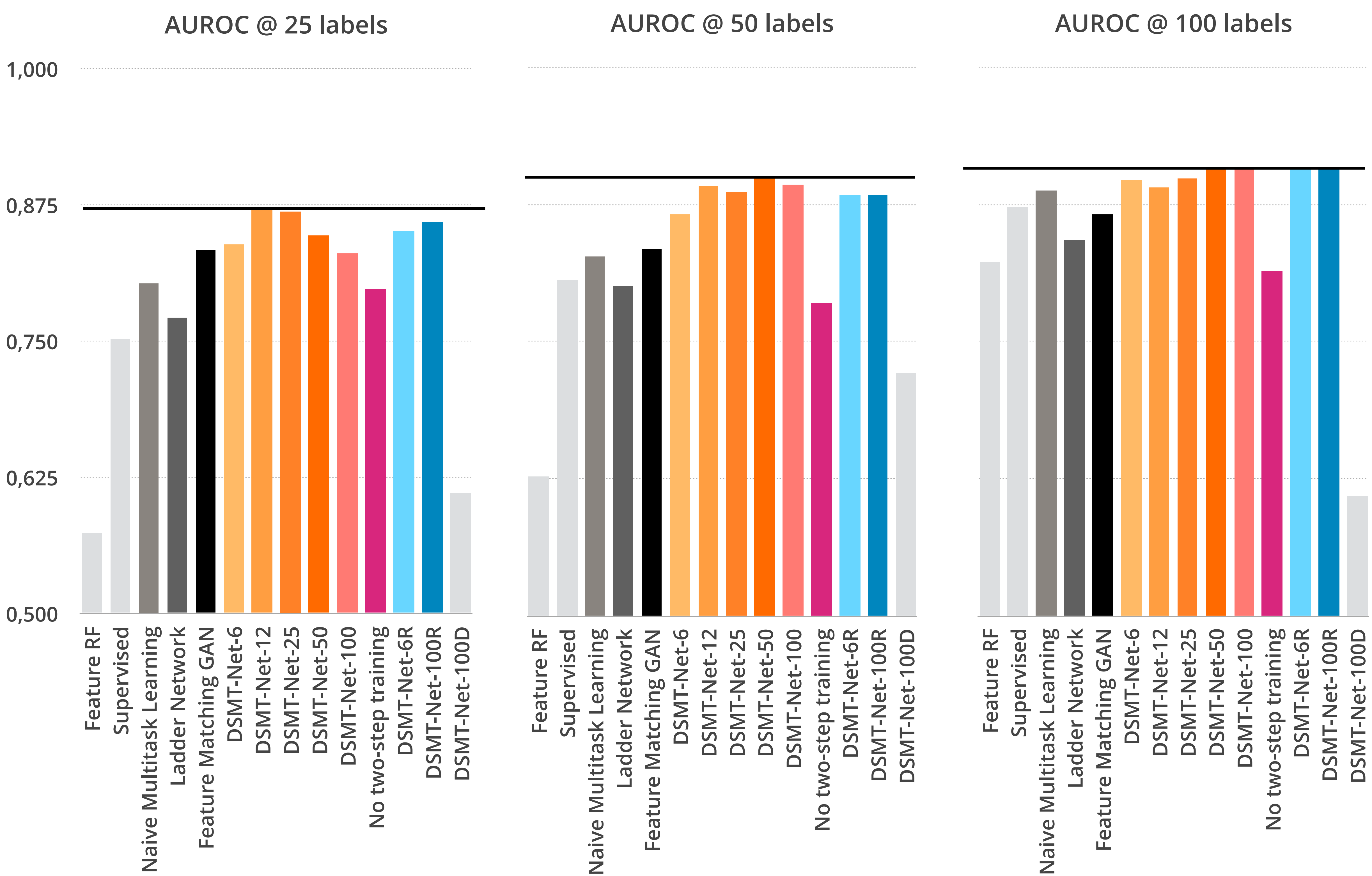

AUROC @ 12 labels
DSMT-Nets (importance)
1,000
0,875
0,750
0,625
0,500
Feature RF
Supervised
Naive Multitask Learning
Ladder Network
Feature Matching GAN
DSMT-Net-6
DSMT-Net-12
DSMT-Net-25
DSMT-Net-50
DSMT-Net-100
No two-step training
DSMT-Net-6R
DSMT-Net-100R
DSMT-Net-100D

Results

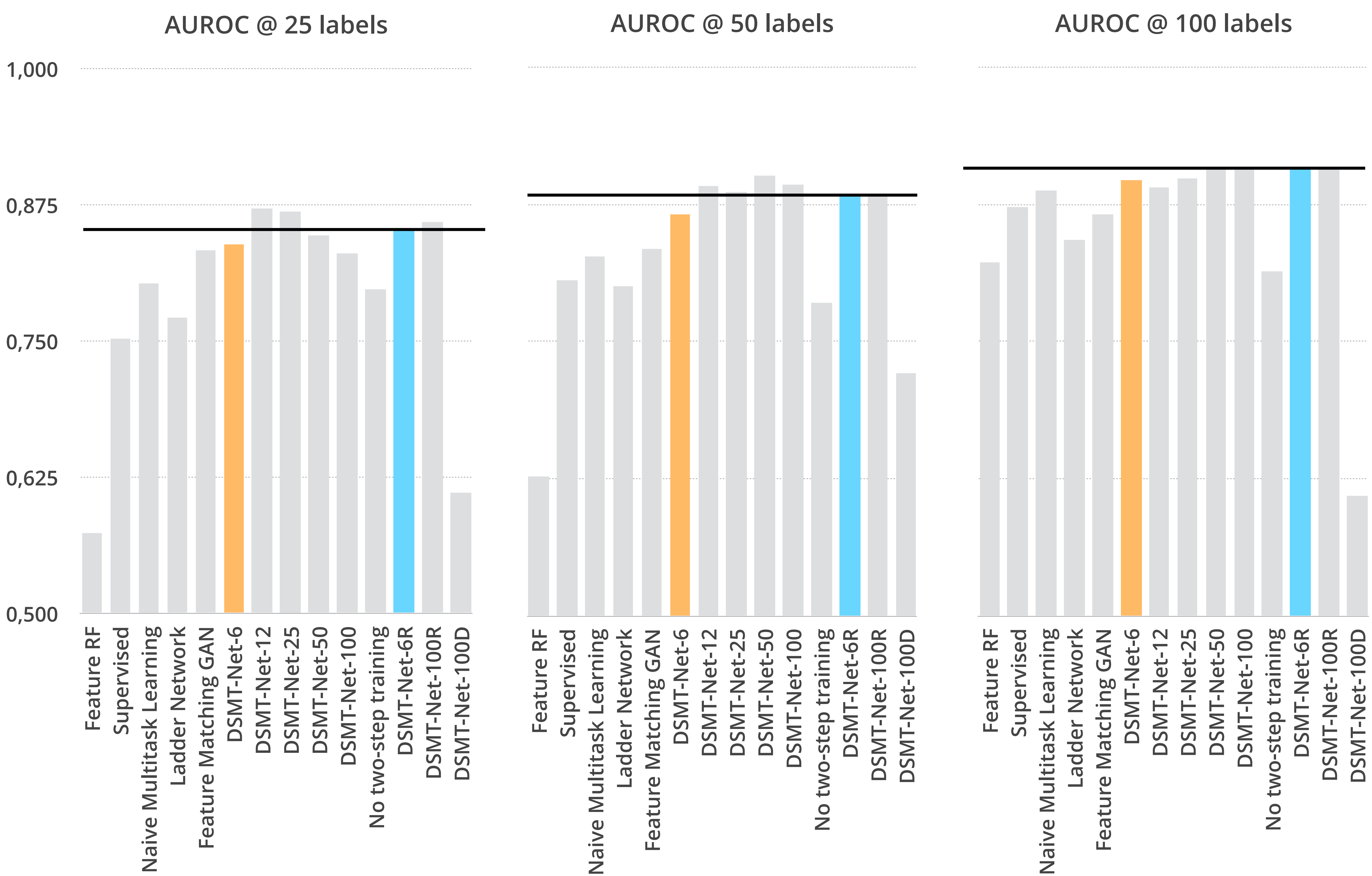


AUROC @ 12 labels
1,000
0,875
0,750
0,625
0,500
DSMT-Nets (R + D)
Feature RF
Supervised
Naive Multitask Learning
Ladder Network
Feature Matching GAN
DSMT-Net-6
DSMT-Net-12
DSMT-Net-25
DSMT-Net-50
DSMT-Net-100
No two-step training
DSMT-Net-6R
DSMT-Net-100R
DSMT-Net-100D

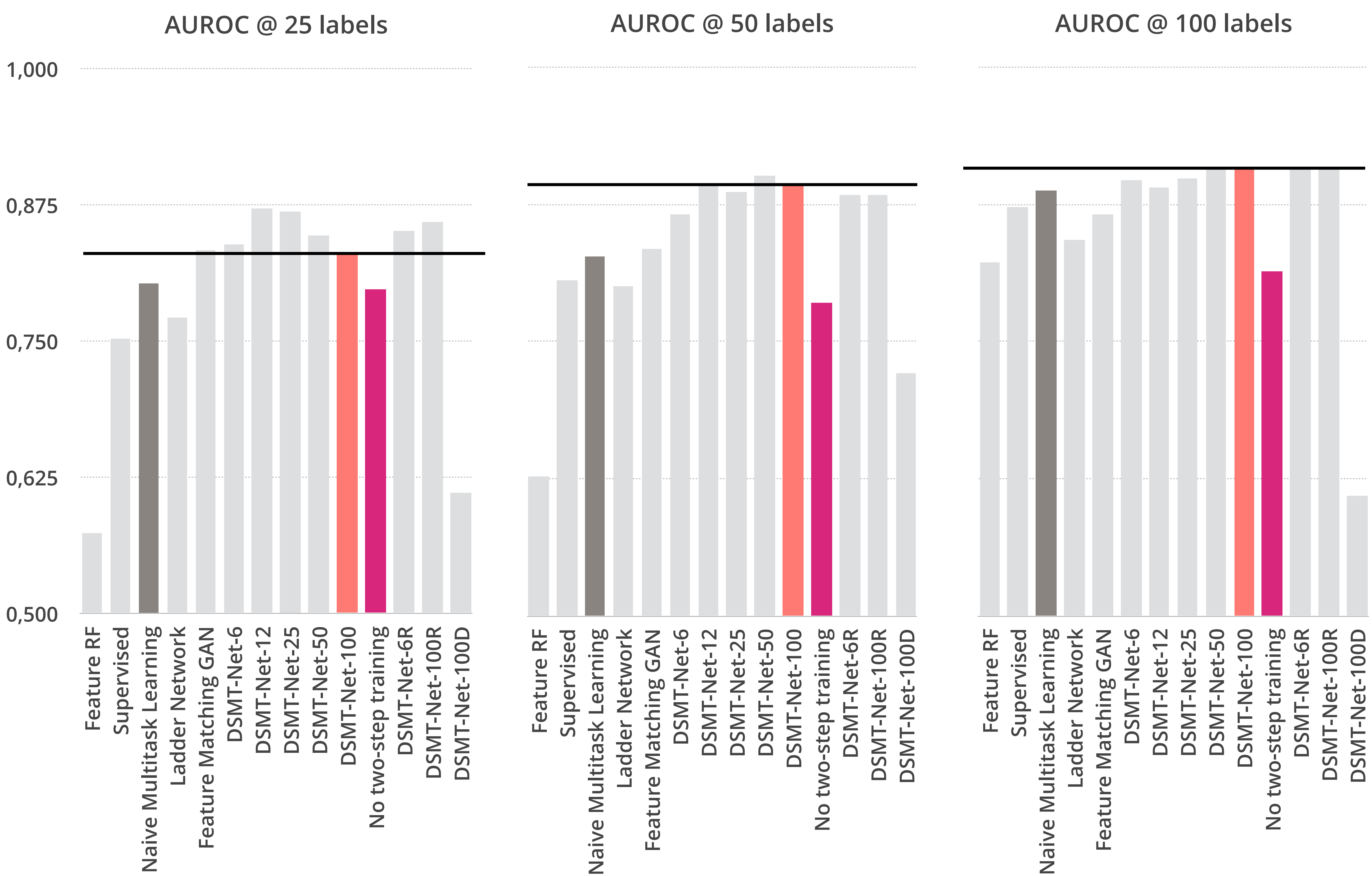
DSMT-Nets outperform existing SSL methods



Random outperforms Importance Selection



Preventing Adverse Gradient Interactions Is Key



Conclusion

Conclusion

- We present an approach to semi-supervised learning that …
  - ✔ **automatically** selects a large set of auxiliary tasks from multivariate time series
  - ✔ scales to **hundreds of auxiliary tasks** in a single neural network
  - ✔ **combats adverse gradient interactions** between tasks
- We confirm that **adverse gradient interactions** and **auxiliary task diversity** are key in multitask learning.
- We make good progress on a **clinically important** task.

Eidgenössische Technische Hochschule Zürich
Swiss Federal Institute of Technology Zurich

ICU Cockpit
artificial intelligence for neurointensive care

Questions?

Patrick Schwab

@schwabpa

patrick.schwab@hest.ethz.ch
Institute for Robotics and Intelligent Systems
ETH Zurich

Find out more at the poster session (#108, 18.15), and in the paper:
Schwab, P., Keller, E., Muroi, C., Mack, D. J., Strässle, C., and Karlen, W. (2018).
Not to Cry Wolf: Distantly Supervised Multitask Learning in Critical Care.

75 NRP Big Data
National Research Programme

University of Zurich UZH