Detection of unmodeled CWs using a Transfer Learning approach

Diego S. Dominguez*, Seiya Sasaoka, Kentaro Somiya and Hirotaka Takahashi

Continuous gravitational waves and neutron stars workshop

KAGRA

東京工業大学
Tokyo Institute of Technology

Somiya Lab.

Hannover, Jun 20th 2024

Today's Outline

Introduction

- Non-axisymmetric neutron stars are expected to emit quasi-monochromatic long-standing continuous waves (CWs).
- Detection of CWs may give valuable insight into open questions about the population of neutron stars and their internal structure.
- Both coherent and semicoherent methods have their limitations.

It would be nice to have a quick look-up tool !

Enter Deep Learning. By using significantly less computational cost, in some applications it has proven to rival matched-filtering sensitivity.

Fig 1. DALL E3's impression of a rotating neutron star

Motivation and objectives

GOAL: Development of a binary classifier to detect CW signals using a Transfer Learning approach

- Instead of development from scratch, we start with a **pretrained** state-of-the-art CNN (ResNeXt-50)
- Train models (Gauss ResNeXt and O3 ResNeXt) to detect CW signals in both Gaussian and real detector noise
- Evaluate the generalization capability of the model

WHY?: Need for quick detection techniques for CW signals using ML.


PHYSICAL REVIEW D **108,** 063021 (2023)

Novel neural-network architecture for continuous gravitational waves

Prasanna M. Joshi[1,2,*] and Reinhard Prix[1,2]
[1]Max Planck Institute for Gravitational Physics (Albert-Einstein-Institute), 30167 Hannover, Germany
[2]Leibniz Universität Hannover, 30167 Hann

(Received 1 May 2023; accepted 14 August 2023; pu

The high computational cost of wide-parameter-space searches
(CWs) significantly limits the achievable sensitivity. This challen
alternative search methods, such as deep neural networks (DNN
convolutional image-classification DNN architectures to all-sky and
for short, one-day search durations, but proved ineffective for longe
paper, we offer a hypothesis for this limitation and propose new desig
of concept, we show that our novel convolutional DNN architecture a
a targeted search (i.e., single sky-position and frequency) in Gaussian
days. We illustrate this performance for two different sky positions a

PHYSICAL REVIEW D **102,** 022005 (2020)

Deep-learning continuous gravitational waves: Multiple detectors and realistic noise

Christoph Dreissigacker* and Reinhard Prix
Max Planck Institute for Gravitational Physics (Albert-Einstein-Institute), D-30167 Hannover, Germany
and Leibniz Universität Hannover, D-30167 Hannover, Germany

ıne 2020; published 6 July 2020)

continuous gravitational waves is limited by
networks (DNNs) can perform all-sky searches
r *et al.*, Phys. Rev. D **100**, 044009 (2019)],
l that could lead to a better overall sensitivity.
g (simulated) strain data from two detectors
sky-position) searches in addition to all-sky
two-detector DNN is about 7% less sensitive
about 51% less sensitive at high frequency
ering (using WEAVE). In the directed case
from about 7%–14% at $f = 20$ Hz to about

Use of Excess Power Method and Convolutional Neural Network in All-Sky Search for Continuous Gravitational Waves

Takahiro S. Yamamoto[1] and Takahiro Tanaka[1,2]
[1]Department of Physics, Kyoto University, Kyoto 606-8502, Japan
[2]Center for Gravitational Physics, Yukawa Institute for Theoretical Physics, Kyoto University, Kyoto 606-8502, Japan
(Dated: March 12, 2021)

The signal of continuous gravitational waves has a longer duration than the observation period. Even if the waveform in the source frame is monochromatic, we will observe the waveform with modulated frequencies due to the motion of the detector. If the source location is unknown, a lot of templates having different sky positions are required to demodulate the frequency, and the required huge computational cost restricts the applicable parameter region of coherent search. In this work, we propose and examine a
coherent search by following-up only t
situation in which only a single-detect
approximated by the stationary Gauss
polarization angle, the inclination an
$\phi_0 = 0$, and they are treated as know
Fourier transform with the re-sample
in some reference direction, 2) the ex
obtained by picking up the amplitud
transform data, and 3) the deep learn
computational cost and the detection
check the validity of the detection pr
for analyzing $O(10^7)$sec strain data.

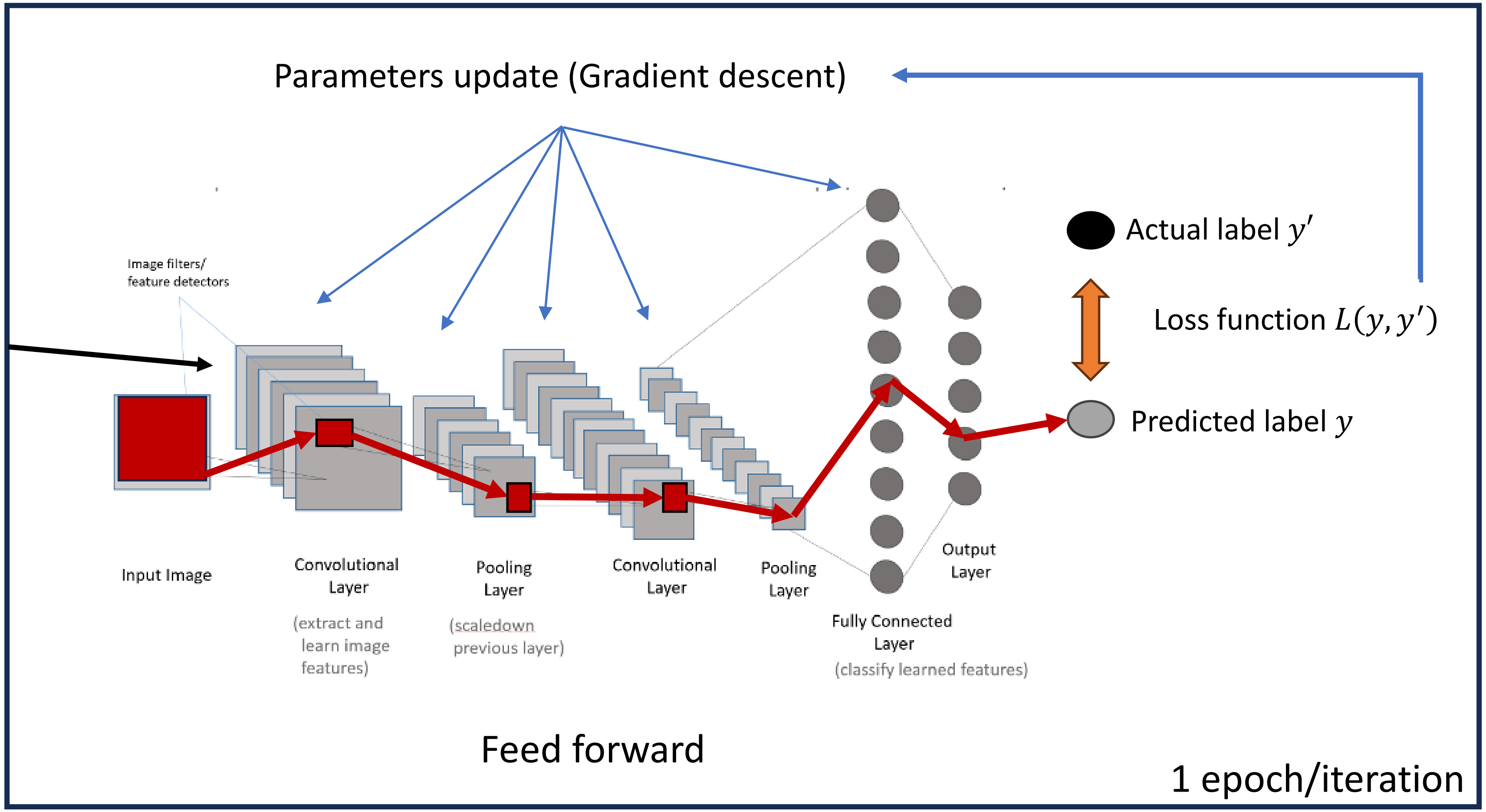
A robust machine learning algorithm to search for continuous gravitational waves.

Joe Bayley,[1] Chris Messenger,[1] and Graham Woan[1]
[1]SUPA, University of Glasgow, Glasgow G12 8QQ, United Kingdom.

Many continuous gravitational wave searches are affected by instrumental spectral lines that could be confused with a continuous astrophysical signal. Several techniques have been developed to limit the effect of these lines by penalising signals that appear in only a single detector. We have developed a general method, using a convolutional neural network, to reduce the impact of instrumental artefacts on searches that use the SOAP algorithm [1]. The method can identify features in corresponding frequency bands of each detector and classify these bands as containing a signal, an instrumental line, or noise. We tested the method against four different data-sets: Gaussian noise with time gaps, data from the final run of Initial LIGO (S6) with signals added, the reference S6 mock data challenge data set [2] and signals injected into data from the second advanced LIGO observing run (O2). Using the S6 mock data challenge data set and at a 1% false alarm probability we showed that at 95% efficiency a fully-automated SOAP search has a sensitivity corresponding to a coherent signal-to-noise ratio of 110, equivalent to a sensitivity depth of 10 $\mathrm{Hz}^{-1/2}$, making this automated search competitive with other searches requiring significantly more computing resources and human intervention.

Deep Learning basics

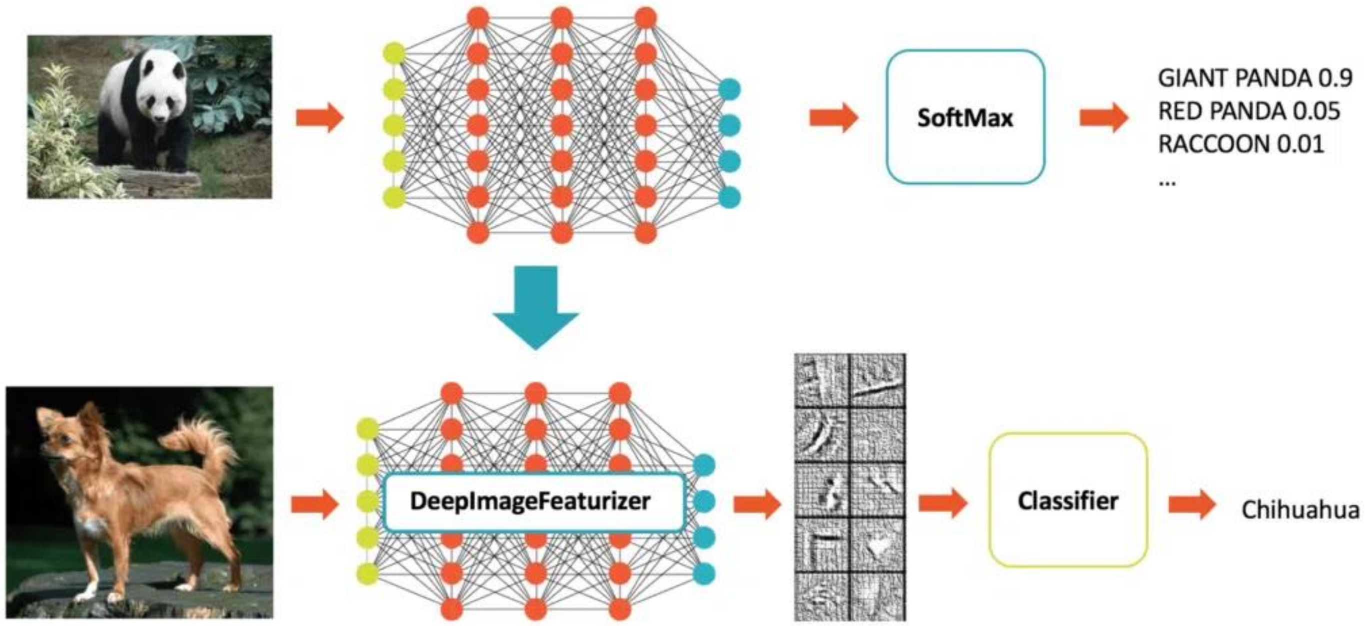
Random initialization of parameters

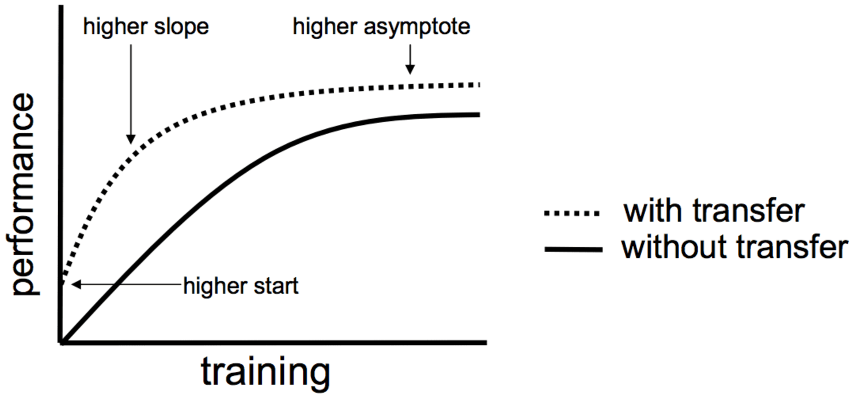


Transfer Learning

We train a base network on a base dataset and task, and then we repurpose the learned features, or transfer them, to a second target network to be trained on a target dataset and task.

- We don't start with randomly initialized parameters
  - It saves time and achieves higher accuracy from epoch 1





ResNext50

- Why not leverage the knowledge of a massive **pre-trained** state of the art CNN ?

  i.e. Inception, **ResNet**, EfficientNet

We chose ResNeXt50 (updated version of ResNet)

- It has residual blocks that preserve information between layers
- 25 million tunable parameters

ImageNet1K

Dataset

1000 classes including:

- Goldfish
- Triceratops
- Cats
- Dogs
- Go-karts
- Furniture

⋮

Input data

Time frequency representation (Complex spectrogram)
SFT length = 30 min

Standard ResNeXt-50 input:

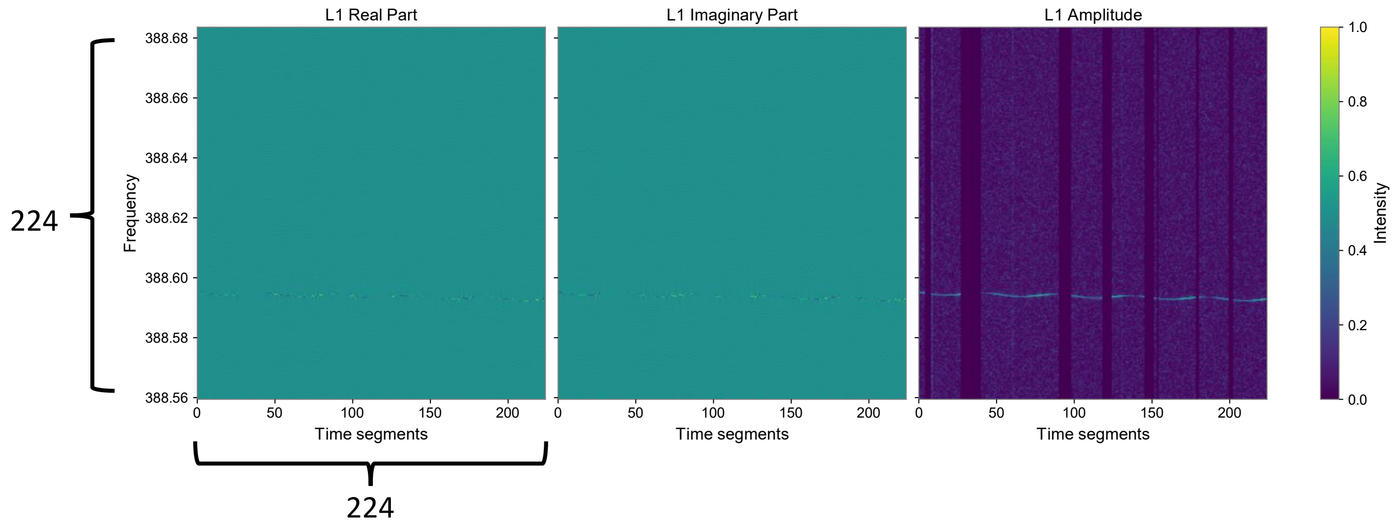
- 3 channels, each 224 px × 224 px

224 frequency bins

$$224 \times \frac{1}{1800}\ \mathrm{Hz} \sim 124\ \mathrm{mHz}$$

224 SFT (each 30 mins)

$$224 \times 30\ \mathrm{min} \sim 4.66\ \mathrm{days}$$



Gaussian noise data (**Gauss ResNeXt**)

- Two detector simulated gaussian noise with realistic time gaps

Real detector data (**O3 ResNeXt**)

- Observation run O3 data from H1 and L1

Parameters prior distributions

	α [rad]	δ [rad]	f [Hz]	$\cos\iota$	ϕ [rad]	ψ [rad]
Lower limit	0	$-\pi/2$	20	-1	0	0
Upper limit	2π	$\pi/2$	1000	1	2π	$\pi/2$
	$\dot{f} = -10^{-9}$ [Hz/s]			band width $= 124$ mHz		

We adopted **Curriculum learning**, network's SNR is the difficulty criterion

$$\rho^2 = \sum_X 4\mathcal{R}\int_0^\infty \frac{\tilde{h}^X(f)\tilde{h}^{X*}(f)}{S_n^X(f)} df,$$

SNR = 500 SNR = 460 … SNR = 40

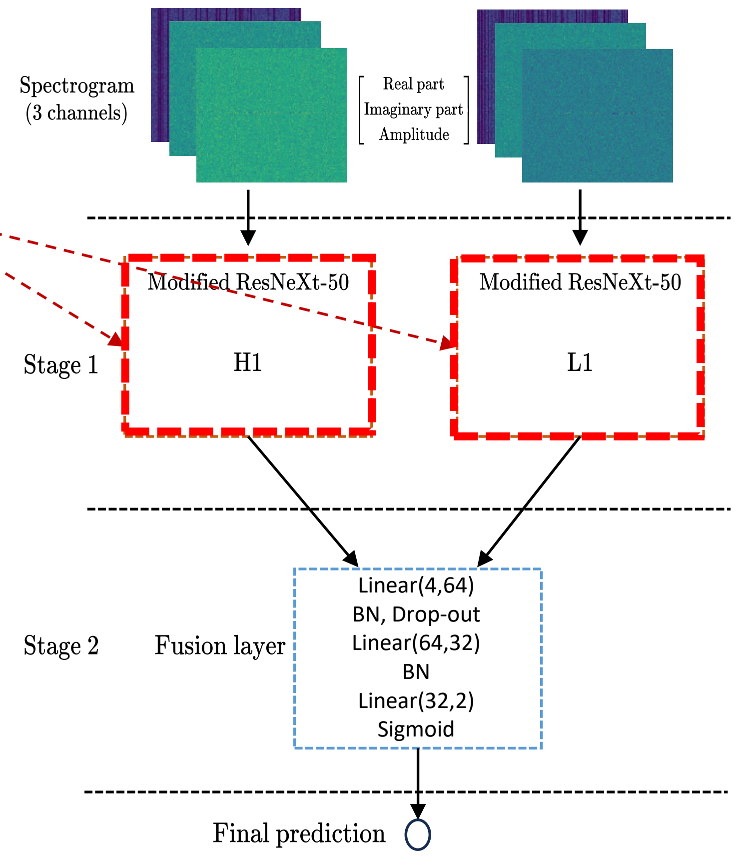
$\sim 3 \times 10^4$ samples

Network architecture

	output	ResNeXt-50 (32 × 4d)	Modified ResNeXt-50
conv0	112×112	$7 \times 7, 64,$ stride 2	$\vdots$
conv1	56×56	$3 \times 3,$ maxpool, stride 2 $\begin{bmatrix} 1 \times 1, 64 \\ 3 \times 3, 128, C = 32 \\ 1 \times 1, 256 \end{bmatrix} \times 3$	$\vdots$ $\vdots$ Frozen parameters
conv2	28×28	$\begin{bmatrix} 1 \times 1, 256 \\ 3 \times 3, 256, C = 32 \\ 1 \times 1, 512 \end{bmatrix} \times 4$	$\vdots$
conv3	14×14	$\begin{bmatrix} 1 \times 1, 512 \\ 3 \times 3, 512, C = 32 \\ 1 \times 1, 1024 \end{bmatrix} \times 6$	$\vdots$
conv4	7×7	$\begin{bmatrix} 1 \times 1, 1024 \\ 3 \times 3, 1024, C = 32 \\ 1 \times 1, 2048 \end{bmatrix} \times 3$	$\vdots$ (Parameter update enabled)
	1×1	global average pool 1000-d fc, softmax	BN 1000-d fc BN, reLU, Drop-out 256-d fc, softmax

25 million trainable parameters

~500.000 trainable parameters



~6 million total trainable parameters

Conventional CNN architecture **lose information**

- Down sampling
- Feature transf. across layers

Residual connections prevent this!

Identity mapping of the input to deeper layers.

This is a **key feature** of the ResNeXt architecture

Training and validation

To ensure effective training:

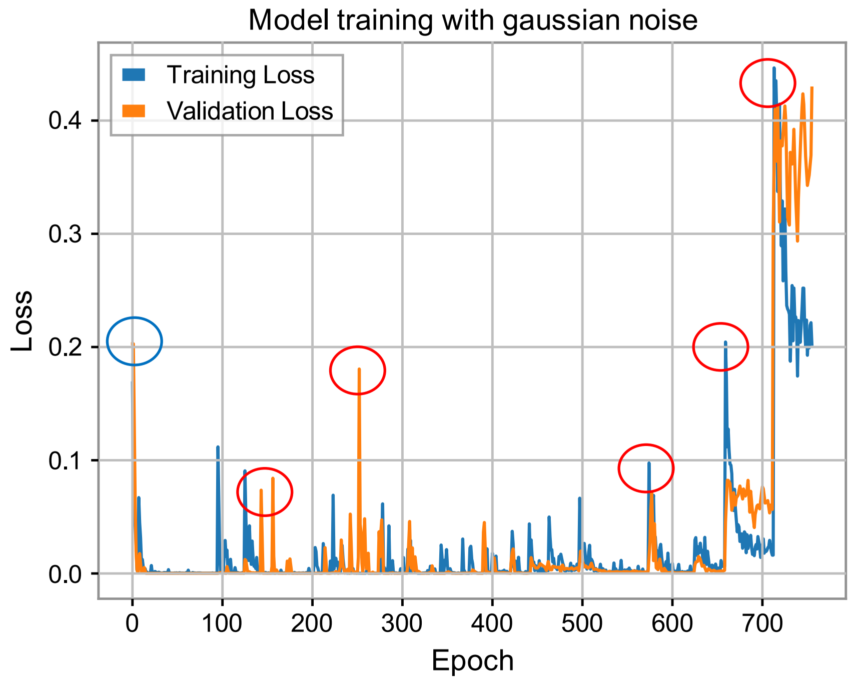
- Hyperparameter tuning with Optuna
- L2 regularization with $\lambda = 10^{-5}$
- Early stopping

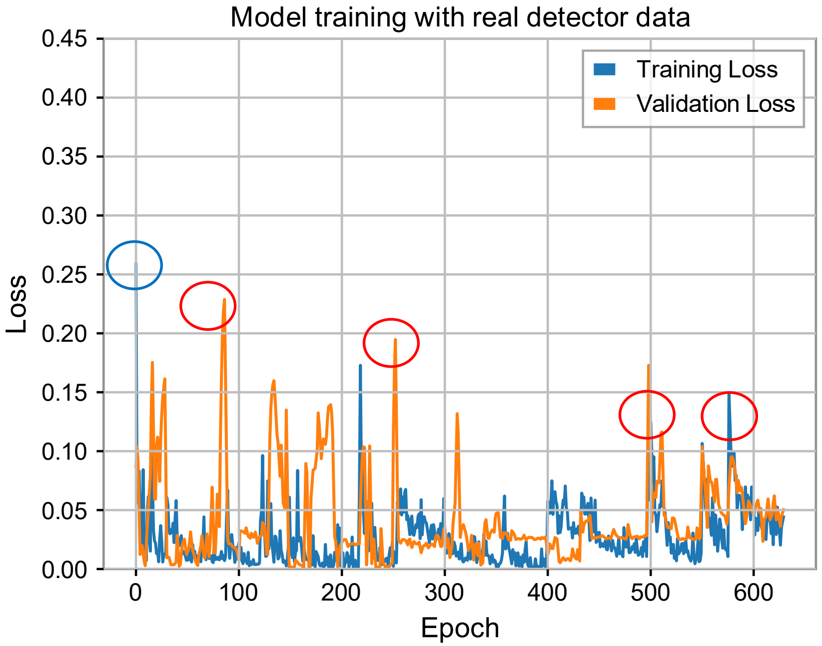
Loss function: cross entropy

True label

$$L_{CE} = -\sum_{i=1}^{N} p_i \log(q_i)$$

Model's prediction



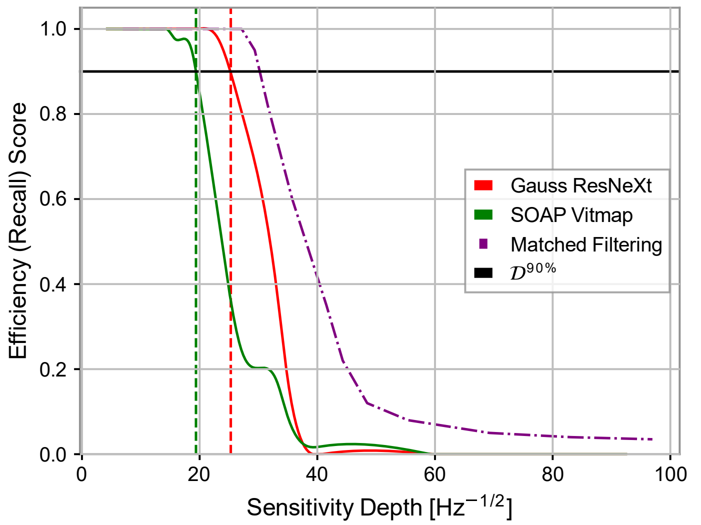


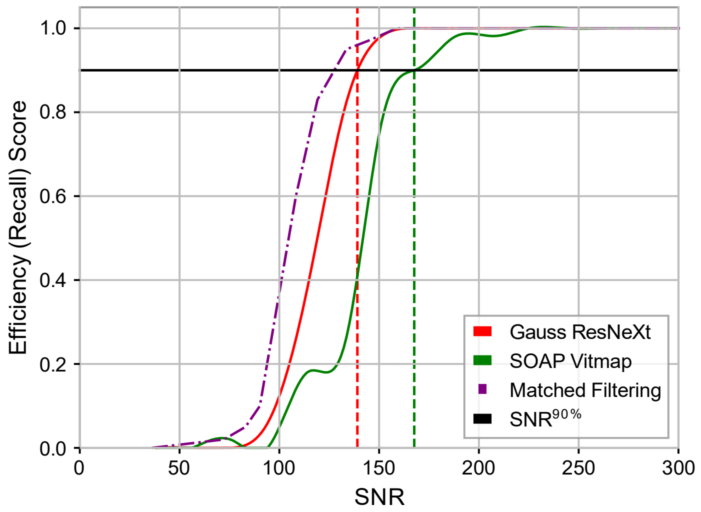
Training time ~ 20.3 hours

Software and Hardware: Pytorch 2.0 implementation / 16 GB NVIDIA TESLA V-100 GPU

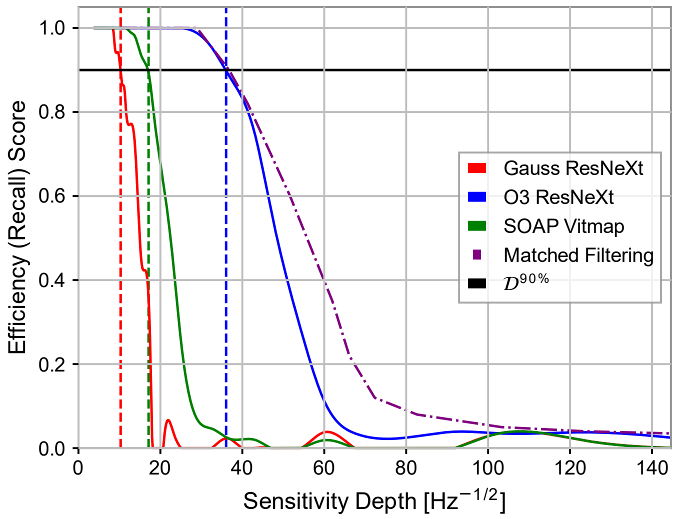
Performance

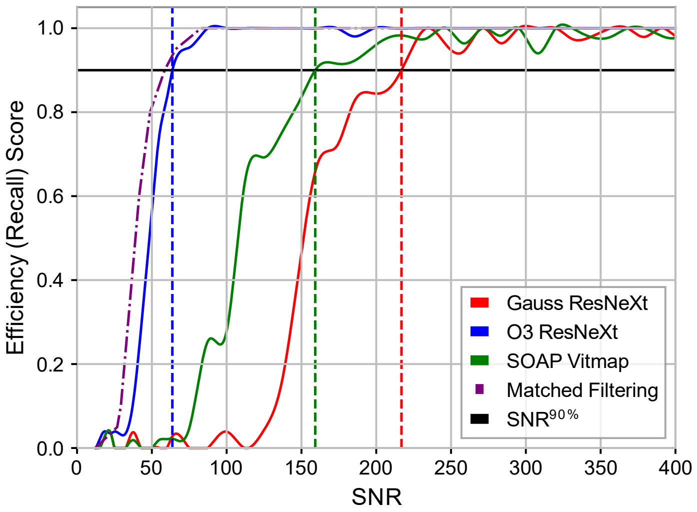
Gaussian Noise





Real detector noise





We compared **Gauss ResNeXt** and **O3 ResNeXt** against the **SOAP CW**[1] pipeline (vitmap) and Matched-Filtering (WEAVE)

Our two metrics are the **SNR** and **sensitivity depth** (at a $p_{\mathrm{fa}} = 1\%$) for which we attain a 90% efficiency of detection

$$\mathcal{D}^{p_{\mathrm{det}}} = \frac{\sqrt{S_n}}{h_0^{p_{\mathrm{det}}}}.$$

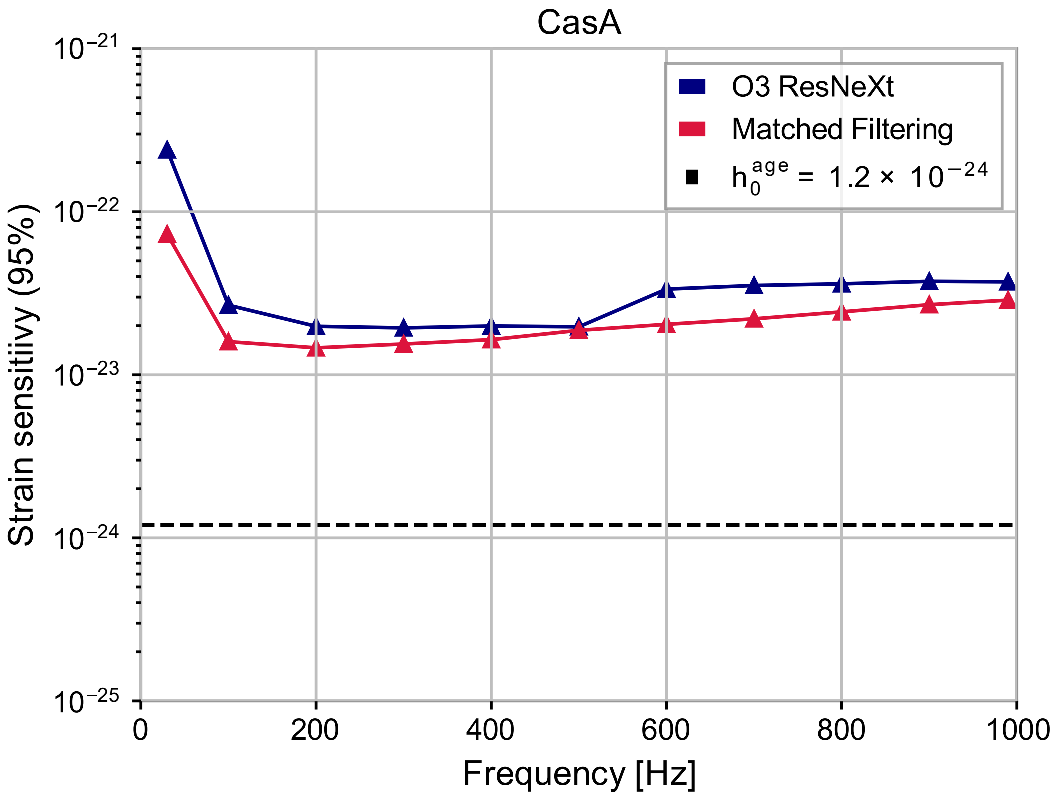
	$\mathcal{D}^{90\%}$		SNR$^{90\%}$	
	Gaussian noise	Real noise	Gaussian noise	Real noise
Matched-Filtering	30.83	36.78	127.54	59.72
Gauss ResNeXt	25.29	9.92	139.09	215.08
O3 ResNeXt	−−	36.04	−−	63.16
SOAP Vitmap	19.37	17.11	167.57	157.90

[1] J. Bayley et al, Phys. Rev. D. 102 083024 (2020)

Detection Efficiency – Targeted search

- As a test benchmark, we compare O3 ResNeXt performance against matched filtering for the Cas A remnant with a strain age $h_0 \sim 10^{-24}$

- Results comparable to matched-filtering



Summary

- Transfer learning is a good alternative to train deeper models in less time
- We developed binary classifier CNNs, **Gauss ResNeXt** and **O3 ResNeXt**, using transfer learning to detect continuous wave (CW) signals in Gaussian and real detector noise.
- Both models achieved detection sensitivities close to matched-filtering for short observation periods (approximately 4.66 days).
- Gauss ResNeXt and O3 ResNeXt showed strong detection capabilities in their respective noise environments but had poor generalization to other noise types.

Future Work: Generalization to longer spectrograms… Bayesian approach or even more DL!