Deep Learning Parameter Estimation of Black Hole Mergers from LIGO Signals

Zain Nasir

Yorktown High School, Yorktown, Indiana

Project Overview

Gravitational wave astronomy provides direct observations of colliding black holes through detectors like LIGO (Laser Interferometer Gravitational-Wave Observatory). By measuring masses, distances, and spins from gravitational wave signals, scientists can reconstruct how binary systems form and evolve. Traditional computational methods demand substantial resources per event, creating detection bottlenecks. Deep learning can accelerate this process by mapping detector signals directly to physical parameters. Multiple neural network architectures exist for gravitational wave analysis, but direct comparisons of their performance is lacking. Another important question is: what do these models actually learn?

Research Questions and Hypotheses

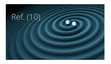
Question: This study compares ResNet and Vision Transformer architectures and asks: a) Which architecture achieves better parameter estimation performance? b) What signal features does each architecture learn? c) Do both models identify physically meaningful patterns?

Hypothesis: Both models will achieve comparable parameter estimation performance. Interpretability analysis will reveal that both architectures focus on the characteristic chirp pattern, while feature extraction mechanisms may differ.

Background Information

What Are Black Holes?

Black holes are regions of spacetime where gravity is so strong that nothing, even light, can escape. Binary black holes are formed where two black holes orbit each other. Over time, these binaries lose energy, causing the orbit to shrink. Then the black holes spiral together and merge.



What Are Gravitational Waves?

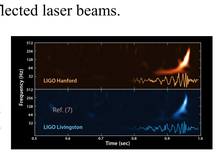
Gravitational waves are ripples in spacetime itself. When massive objects accelerate, they create wrinkle in the fabric of spacetime; these wrinkles then propagate outward as waves traveling at the speed of light, like water ripples produced by a stone thrown into water. The amplitude of gravitational waves is extraordinarily small, making gravitational waves extremely difficult to detect.

LIGO

The Laser Interferometer Gravitational-Wave Observatory (LIGO) consists of two facilities in Hanford, WA and Livingston, LA. Each facility uses laser interferometry to measure spacetime distortions. Two 4-kilometer arms extend perpendicular to each other. Laser light travels down each arm and reflects back.

When a gravitational wave passes through, it stretches spacetime in one direction while compressing it in the perpendicular direction, creating a small difference in the arm lengths that can be measured by comparing the reflected laser beams.

Black Hole Mergers



LIGO first detected gravitational waves in 2015, observing the merger of two black holes approximately 1.3 billion light-years away. By analyzing the gravitational wave signal, scientists can determine the properties of black holes and help answer questions such as how they form.

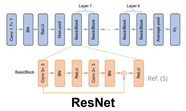
The gravitational wave signal from a merger has a characteristic pattern called a "chirp." As the black holes spiral together, they orbit faster and emit gravitational waves at increasing frequency and amplitude. The signal starts at low frequencies when the black holes are far apart, then sweeps upward in frequency as they approach.

Computational Challenge and Solutions

Current parameter estimation is a very time consuming process. Neural networks offer a faster alternative. By representing signals as time-frequency spectrograms, the training task becomes a problem of image analysis.

ResNet

ResNet processes images through convolutional layers that apply spatial filters to detect patterns. Early layers identify simple features like edges. Deeper layers combine these into complex patterns. The network builds understanding hierarchically from local features to global structure. For spectrograms, ResNet's filters detect frequency patterns and temporal features, building from local frequency bands to overall signal evolution.



ResNet

Vision Transformer (ViT)

Vision Transformer

Vision Transformer divides the input into patches (e.g. 8×8 pixel squares) and uses self-attention instead of convolutions.

What is Attention?

Self-attention computes relationships between all patches simultaneously. For each patch, the model calculates attention weights, numerical values indicating how much to focus on every other patch. High weight means that patch is important.

Multi-Head Attention

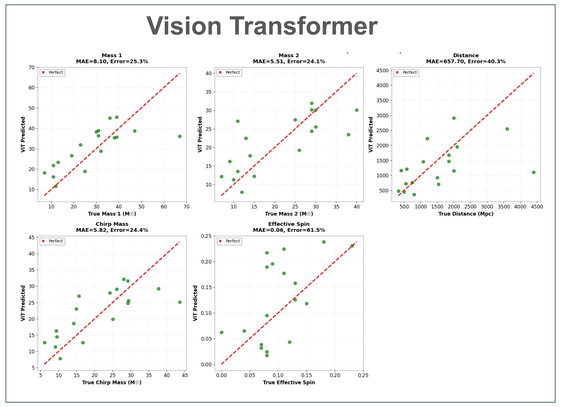
Multiple sets of attention weights computed in parallel, each capturing different relationships. The model has four transformer blocks, each applying multi-head attention followed by a feedforward network. For spectrograms, attention identifies which time-frequency regions are related. The model processes the entire signal evolution simultaneously.

References

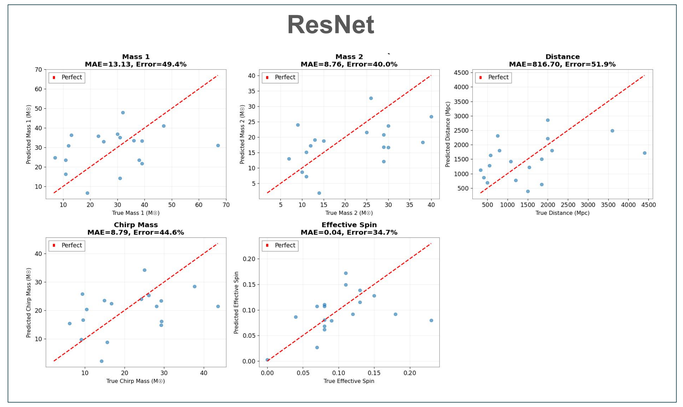
1. "Gravitational Waves, as Einstein Predicted." *LIGO Lab | Caltech*, www.ligo.caltech.edu/image/ligo20160211a.
2. "Gravitational Waves from a Binary Black Hole Merger Observed by LIGO and Virgo." *LIGO Lab | Caltech*, www.ligo.caltech.edu/news/ligo20170927.
3. "GWTC." *Gwosc.org*, gwosc.org/eventapi/html/GWTC/.
4. Hampson, Michelle. "Sensor Could Help Detect Gravity Waves in Orbit." *IEEE Spectrum*, July 2024, spectrum.ieee.org/gravitational-wave-detector.
5. Lerner, Louise. "What Is a Black Hole?" *News.uchicago.edu*, 13 Oct. 2022, news.uchicago.edu/explainer/black-holes-explained.
6. LIGO Caltech. "What Is an Interferometer?" *LIGO Lab | Caltech*, 2019, www.ligo.caltech.edu/page/what-is-interferometer.
7. Sutter, Paul. "What Happens When Black Holes Merge?" *Space.com*, Space, 21 Oct. w024, www.space.com/what-happens-when-black-holes-merge.
8. Dosovitskiy, Alexey, et al. "An image is worth 16x16 words: Transformers for image recognition at scale." International Conference on Learning Representations (2021).
9. He, Kaiming, et al. "Deep residual learning for image recognition." Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (2016): 770-778.
10. Iwanaga, Hibiki, Mahoro Matsuyama, and Yousuke Itoh. "Enhancing the reliability of machine learning for gravitational wave parameter estimation with attention-based models." Physical Review D 112.8 (2025): 083020.
11. Raschka, Sebastian, Yuxi (Hayden) Liu, and Vahid Mirjalili. "Machine Learning with PyTorch and Scikit-Learn." Packt Publishing (2022).
12. Ruan, Wen-Hong, et al. "Rapid search for massive black hole binary coalescences using deep learning." Physics Letters B 841 (2023): 137904.
13. Schmidt, Stefano, et al. "Machine learning gravitational waves from binary black hole mergers." Physical Review D 103.4 (2021): 043020.

Results: Parameter Estimation

Vision Transformer



ResNet



Performance Comparison

Primary mass (m_1):
- Vision Transformer: 25.3% median error, 8.10 M⊙ mean absolute error
- ResNet: 49.4% median error, 13.13 M⊙ mean absolute error

Secondary mass (m_2):
- Vision Transformer: 24.1% median error, 5.51 M⊙ mean absolute error
- ResNet: 40.0% median error, 8.76 M⊙ mean absolute error

Luminosity distance:
- Vision Transformer: 40.3% median error, 657.70 Mpc mean absolute error
- ResNet: 51.9% median error, 816.70 Mpc mean absolute error

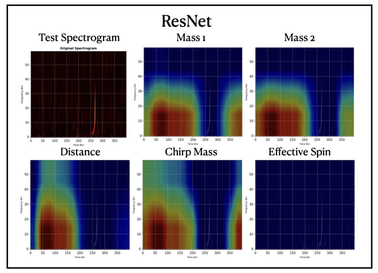
Chirp mass:
- Vision Transformer: 24.4% median error, 5.82 M⊙ mean absolute error
- ResNet: 44.6% median error, 8.79 M⊙ mean absolute error

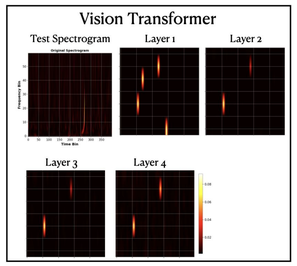
Effective spin (chi_eff):
- Vision Transformer: 61.5% median error, 0.06 mean absolute error
- ResNet: 34.7% median error, 0.04 mean absolute error

Synopsis

- Overall, in terms of median error, the Vision Transformer achieved 35.12 ± 7.26% compared to ResNet's 44.12 ± 3.13%.
- While ResNet exhibits a higher mean error, the substantial overlap in error bounds (ViT: 27.9–42.4%, ResNet: 41.0–47.3%) indicates no statistically significant performance difference.

Results: Interpretability Analyses





Interpretability Analyses

Vision Transformer Attention Evolution

- Attention maps across the transformer layers show progressive spatial refinement:
  - Layer 1 exhibits distributed attention across multiple time bins with vertical structure.
  - Layers 2 and 3 show increasing concentration while maintaining vertical organization.
  - Layer 4 displays sharp vertical features at specific temporal locations, representing the final features used for parameter prediction.
- The model focuses on discrete time bins extending across frequency ranges. The temporal localization strategy identifies when signal characteristics are most informative.

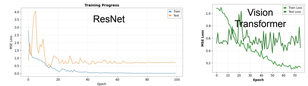
ResNet Grad-CAM Spatial Patterns

- Grad-CAM heatmaps reveal parameter-specific importance patterns:
  - Primary mass: Horizontal band organization; high importance (red) in 0–25 Hz extending across time bins, with moderate importance (yellow/orange) to 35 Hz.
  - Secondary mass: Similar horizontal structure, 0-20 Hz high importance extending to ~40 Hz mid-frequencies.
  - Distance: Asymmetric pattern with high importance concentrated in early time bins (0–50 ms) at low frequencies (0–30 Hz).
  - Chirp mass: Broad frequency coverage (0–40 Hz) with horizontal band structure across time.
  - Effective spin: Minimal activation throughout spectrogram.
- All parameters emphasize lower frequencies (0–30 Hz) with horizontal organization.

Comparative Interpretability Analysis

- Vision Transformer and ResNet employ different spatial organizations:
  - ViT: Vertical features indicating temporal localization
  - ResNet: Horizontal features indicating frequency band selection
- Both architectures focus on signal-containing regions (mid-to-late time, lower frequencies) while avoiding noise-dominated areas.
- Both approaches successfully identify gravitational wave signal regions, validating that learned features align with expected signal locations rather than noise or artifacts.

Training and Validation



Concluding Remarks

Conclusions

- Both Vision Transformer and ResNet achieved comparable performance.
- Interpretability analysis revealed distinct feature extraction strategies: Vision Transformer attention organized as vertical features (temporal localization), ResNet Grad-CAM as horizontal features (frequency band selection).
- Both architectures learned physically meaningful features:
  - Vision Transformer focused on chirp evolution.
  - ResNet emphasized lower frequencies.
  - Both avoided noise-dominated regions.
- Vision Transformer achieved comparable accuracy with 22x fewer parameters (0.49M vs 11M).
- Both models successfully learned from limited number of real LIGO events, achieving errors within a factor of 2 of published benchmarks using thousands of simulated signals.

Future Directions

- Explore hybrid architectures combining convolutional and transformer components, and multi-detector models.
- As additional merger detections accumulate, expanding the training set will assess how architectures scale with data.
- Uncertainty quantification to predict confidence intervals.
- Testing on neutron star mergers to generalize across source types.

Limitations

- Small test set limits statistical power for architecture comparison.
- Models predict only five parameters.
- Each detector's data were treated independently though the same event is observed by both LIGO facilities.

Methodology

Data Collection and Processing

Gravitational wave data from 54 confirmed black hole merger events were obtained from the Gravitational Wave Open Science Center (GWOSC). Raw detector signals were processed using GWOSC analysis software.

Data were downsampled from 4096 Hz to 2048 Hz. Whitening was applied to flatten the noise spectrum (making all frequencies comparable intensity). A bandpass filter retained frequencies between 20 and 500 Hz. Time-series signal data were converted to spectrograms using the Q-transform, which provides time-frequency representations showing how signal frequency changes over time as black holes spiral together.

Black Hole Parameters

Five physical parameters were selected as prediction targets: Primary black hole mass (m1); Secondary black hole mass (m2) in solar masses; Luminosity distance (Mpc); Chirp mass; Effective spin parameter.

Neural Network Architectures

The Vision Transformer (ViT: 0.49M parameters) divides input spectrograms into non-overlapping patches of 8×8 pixels with each patch projected to a 64-dimensional embedding space. Four transformer encoder blocks process the embedded patches, each containing multi-head self-attention with 4 heads and feedforward networks with hidden dimension 256. Layer normalization along with dropout of 0.15 were applied for regularization. A final layer maps representations to 5 output parameters.

The ResNet (11M parameters) implements a convolutional architecture based on ResNet-18. The architecture begins with a 7×7 convolutional layer with 64 output channels, followed by batch normalization, ReLU activation, and 3×3 max pooling. Each basic block consists of two 3×3 convolutional layers with batch normalization, ReLU activation, and skip connections, and a final linear layer mapping to 5 outputs.

Training Protocol

Training used identical data and similar protocols: combined loss function (70% mean squared error and 30% mean absolute error), AdamW optimizer with learning rate 0.001 and weight decay 0.01, batch size 2, and up to 150 epochs.

Interpretability Analysis

After training, interpretability techniques examined what each architecture learned. AI models must learn physically meaningful features rather than spurious correlations with noise or instrumental artifacts. Interpretability analysis validates that models focus on genuine gravitational wave signals.

Vision Transformer: Attention weights from all four transformer layers were extracted during inference. In the Vision Transformer, each spectrogram patch attends to all other patches through self-attention, computing numerical weights that indicates relationship strength. High attention weight means a patch is important for understanding another patch. These weights are learned during training, not pre-programmed.

The model has 4 attention heads per layer, each potentially focusing on different relationships. Weights from the 4 heads were averaged to produce layer-specific attention maps showing where the model focused when making predictions. This reveals the model's decision process: Layer 1 shows broad exploration, while Layer 4 shows concentrated focus on specific features used for final predictions. Attention maps were visualized as heatmaps. Bright regions indicate high attention (the model focuses here), dark regions indicate low attention (ignored).

ResNet: Gradient-weighted Class Activation Mapping (Grad-CAM) was applied to visualize which spectrogram regions influenced ResNet predictions. This technique asks: "If this region is modified, how much would the prediction change?" Regions strongly affecting predictions are important; regions barely changing predictions are less relevant. Grad-CAM computes gradients of the predicted parameters with respect to the final convolutional layer (layer 4) feature maps. Gradients measure sensitivity, how output changes when a spatial location changes. High gradient magnitude indicates high importance. These gradients are averaged to obtain importance weights.

The weighted combination of feature maps produces a heatmap. Hot regions indicate high importance (strongly influence predictions). Cold regions indicate low importance (minimal influence on predictions). Grad-CAM was computed separately for each of the five predicted parameters, revealing whether different parameters use different spectrogram regions.

Comparison and Validation: Both analyses were performed on the same test sample (sample 5 is shown) to enable direct comparison. The visualizations could reveal different organizational strategies. By examining whether both methods highlight signal-containing regions while avoiding noise-dominated areas, we validate that both architectures learned to identify and extract information from genuine gravitational wave signals rather than artifacts. The spatial patterns are compared to the expected chirp pattern to verify physical alignment.

Compute Resources, Implementation and Analyses

Model training was performed on Google Colab using NVIDIA Tesla L4 GPUs. Both models were trained in less than 60 minutes. Both models were implemented using PyTorch 2.0+. The implementations followed free online resources. No pre-trained weights were used. NumPy, Matplotlib were used for data handling and visualization. Model performance was evaluated on the test set. Results are reported as median relative error percentages and mean absolute errors.

Spectrograms: Spectrograms show frequency versus time (horizontal axis), with color indicating signal strength. The characteristic chirp pattern with the upward frequency sweep shows spiralling black holes merging together. The spectrograms are input images processed by neural networks.