SOCIETY FOR
NATURAL
SCIENCES
STUDENT JOURNAL

Low Galaxy Radii Dark Matter Predictions: A Hybrid Attention based Learning and Classical Modeling approach.

Danny Cooper

Natural Sciences, Faculty of Science, University of East Anglia

Supervisor Dr Robert Ferdman

1 Abstract

The discrepancy between observed galaxy rotation curves and calculations via Newtonian dynamics has long evidenced the presence of dark matter. Classically represented dark matter halo models, such as the Navarro-Frenk-White (NFW), Einasto, and DC14 profiles, are frequently used to infer the distribution of this unseen mass. Whilst effective at large galaxy radii, these models often fail to accurately represent the shape of rotation curves at central regions of a galaxy. This failure arises from fixed parametric forms within the models that cannot account for the complex interplay between baryonic physics and dark matter.

This literature review explores the embedding of transformer-based machine learning (ML) models within classical models to improve the representation of dark matter inferences. Specifically, transformers trained on galaxy feature datasets learn to predict the parameters retained within these analytic profiles; optimising their fit via learnable weights. By introducing self-attention mechanisms, transformer models identify contextual inter dependencies between observational inputs—such as HI linewidths, stellar densities, and scale radii—and output refined parameter estimates that resolve low-radius modeling errors.

2 Introduction

2.1 The Dark Matter Challenge in Galactic Dynamics

Dark matter is hypothesised to account for 27% of the universe's mass-energy content, yet its composition remains undefined. The induced gravitational influence is most clearly observed in the dynamics of galaxies, particularly through their rotation curves. Rotation curves, which plot the orbital velocity of stars and gas as a function of galactic radius, provide indirect evidence of dark matter, as visible mass alone cannot account for the observed velocities at given radii.

Since the foundational work of Rubin and Ford in the 1970s [1], astronomers have identified a persistent flatness of rotation curves extending considerably further than the optical edge of galaxies. Classical Newtonian calculations, based solely on luminous matter, predict a declining curve akin to Keplerian motion. In reality, astronomical observations show that rotational velocities remain constant or even increase with galaxy radii, suggesting the presence of a more massive, non-luminous component — dark matter — surrounding the galaxy.

2.2 Classical Modeling Approaches

To provide reasoning for these observations, theoretical frameworks introduce multiple parameterised dark matter halo models. The most common profiles—Navarro-Frenk-White (NFW)[2], Einasto[3, 4], and DC14[5]—define functional forms for the halo density with repect to radius. These models are typically fit to observational data by optimising a small number of parameters, such as central density or scale radius, under various assumptions such as, spherical symmetry and static equilibrium.

While these profiles perform successfully at larger galaxy radii, they often diverge from observations at inner regions (typically < 5 kpc) of galaxies, as W. J. G. de Blok (2010) and K. A. Oman et al (2015) note[6, 7]. At these radii, discrepancies between model predictions and observed velocities become more pronounced. This is most likely due, in part, to complex baryonic processes—such as stellar feedback, disk instabilities, and gas inflows—that are not captured by the parameterisations used as within classical models. Furthermore, degeneracies between dark matter halo parameters and the baryonic mass-to-light ratio further complicate the inference of mass distributions as described by S. Courteau et al. (2014)[8].

2.3 Motivation for Machine Learning Integration

Recent construction of astronomical surveys, such as SPARC [9] and SDSS [10], have generated large, broad scope data sets capturing multi-wavelength galaxy properties. These data sets motivate the necessity of modeling methods that can actively accommodate the complexity of galaxy dynamics. Machine learning (ML) architectures, specifically deep learning models, have already demonstrated significant benefit across astrophysical domains: for example, work by A. D'Isanto and K. Polsterer, (2018) in photometric redshift estimation [11], galaxy morphology classification from S. Dielema, et al (2015) [12], and gravitational wave signal detection displayed by D. George and E. A. Huerta (2018),[13].

Transformer-based architectures, originally implemented for natural language processing as showcased by A. Vaswani et al (2017)[14], offer an especially promising capability through multi-head attention. Their self-attention mechanisms enable the representation of contextual relationships across all input features without assuming a temporal relationship. In astrophysical applications, W. Luo, et al (2023) explore the broad use of transformers for generalist applications within astronomy research [15].

This literature review focuses on the integration of transformers within the classical dark matter halo predictive approach. Rather than aiming to discard and replace physical models, transformers are employed to predict optimal halo parameters from galaxy features, thereby refining classical profiles using galaxy features based on data-driven insights. This hybrid methodology leverages both the interpretability of physical models and the predictive power of attention-based ML, offering a new path toward resolving persistent discrepancies in rotation curve modeling.

3 Galaxy Rotation Curves and Dark Matter

3.1 Foundations of Observational Evidence

The rotation curve of a galaxy describes how the orbital velocity of stars and gas vary as a function of distance from a galaxy's center. According to Newtonian dynamics, a galaxy dominated by its luminous mass should exhibit a rotation curve that peaks at the center with a decline as radius increases, following a $v(r) \propto r^{-1/2}$ relationship typical of Keplerian motion. However, spectroscopic measurements by Rubin and Ford in galaxies such as M31 revealed a contradictory finding: rotation curves remained flat or tended to slightly rise well beyond the visible disk [1].

This finding has since been confirmed across numerous spiral galaxies, as documented in large datasets such as the SPARC database [9]. These profiles indicate the presence of an extended mass component that does not emit light—dark matter—encompassing the visible galaxy.

3.2 Classical Dark Matter Halo Models

To reconcile observations with theory, various halo density profiles have been proposed. The most widely used include:

- **Navarro-Frenk-White (NFW) Profile** [2]:

  $$\rho_{\mathrm{NFW}}(r) = \frac{\rho_0}{\left(\frac{r}{r_s}\right)\left(1 + \frac{r}{r_s}\right)^2}$$

  This profile arises from ΛCDM simulations and assumes a cuspy inner density structure.

- **Einasto Profile**[3, 4]:

  $$\rho_{\mathrm{Ein}}(r) = \rho_e \exp\left[-\frac{2}{\alpha}\left(\left(\frac{r}{r_e}\right)^{\alpha} - 1\right)\right]$$

  Offering greater flexibility, the Einasto profile can accommodate varying central steepness.

- **DC14 Model** [5]An empirical model incorporating baryonic feedback effects, allowing core-like structures in dwarf galaxies:

  $$\rho(r) = \frac{\rho_s}{\left(\frac{r}{r_s}\right)^{\gamma}\left[1 + \left(\frac{r}{r_s}\right)^{\alpha}\right]^{(\beta-\gamma)/\alpha}}$$

  where the inner slope γ, the outer slope β, and the sharpness parameter α depend on the stellar-to-halo mass ratio as:

  $$\gamma = 0.5 + \log_{10}\left(\frac{M_\star}{M_{\mathrm{halo}}}\right)$$

  $$\beta = 3.0 + 0.25 \times \log_{10}\left(\frac{M_\star}{M_{\mathrm{halo}}}\right)$$

  These relations adjust the profile shape to account for baryonic effects like stellar feedback.

These models are typically fit using optimisation techniques such as χ^2 minimisation or Bayesian inference. Parameters like the scale radius r_s and characteristic density ρ_0 are adjusted to match observed curves.

3.3 Worked Example: Model Failure at Low Radii

Considering a disc galaxy with a visible mass of $M_d = 1 \times 10^9\, M_\odot$ and a disc scale length of $R_d = 1.5$ kpc highlights the discrepancies between observed and predicted rotational velocities at low radii.

The contribution to the rotational velocity from the stellar disc can be computed using the Freeman exponential disk model, K. C. Freeman, (1970) [17]:

$$v_{\mathrm{disk}}^2(r) = \frac{GM_d}{2R_d} y^2 \left[I_0(y)K_0(y) - I_1(y)K_1(y)\right], \quad \text{where} \quad y = \frac{r}{2R_d},$$

and I_n, K_n are modified Bessel functions. At $r = 0.5$ kpc, this yields:

$$v_{\mathrm{disk}}(0.5\,\mathrm{kpc}) \approx 7.3\,\mathrm{km/s}.$$

Observational data from SPARC [9] for similar dwarf galaxies typically show:

$$v_{\text{obs}}(0.5\,\text{kpc}) \approx 25\,\text{km/s}.$$

Adding a classical NFW halo with parameters $\rho_0 = 0.05\,M_\odot\,\text{pc}^{-3}$ and $r_s = 5$ kpc, the predicted dark matter contribution is calculated as:

$$M_{\text{NFW}}(r) = 4\pi\rho_0 r_s^3 \left[\ln\left(1 + \frac{r}{r_s}\right) - \frac{r/r_s}{1 + r/r_s}\right],$$

which yields:

$$M_{\text{NFW}}(0.5\,\text{kpc}) \approx 3 \times 10^8\,M_\odot, \quad \Rightarrow \quad v_{\text{NFW}}(0.5\,\text{kpc}) \approx 52\,\text{km/s}.$$

Total predicted velocity:

$$v_{\text{total,NFW}}(0.5\,\text{kpc}) = \sqrt{(7.3)^2 + (52)^2} \approx 52.5\,\text{km/s},$$

which significantly overestimates the observed 25 km/s velocity.

Using a DC14 halo, which reduces central mass density via baryonic feedback effects, suppose:

$$M_{\text{DC14}}(0.5\,\text{kpc}) \approx 5 \times 10^7\,M_\odot \quad \Rightarrow \quad v_{\text{DC14}}(0.5\,\text{kpc}) \approx 21\,\text{km/s},$$

then total:

$$v_{\text{total,DC14}}(0.5\,\text{kpc}) = \sqrt{(7.3)^2 + (21)^2} \approx 22.2\,\text{km/s},$$

which more closely matches the observed value.

This example demonstrates that classical NFW profiles, often significantly overestimate, while other models underestimate, velocities in central regions of low-mass galaxies, reinforcing the need for more dynamic modeling methods that can adapt to galaxy features responsible for halo structures.

4 Machine Learning Applications in Astrophysics

Machine learning (ML) has increasingly been incorporated into astrophysics, primarily used in the analysis of large, complex datasets that traditional modeling techniques struggle to computationally manage. Through outlining key developments in the application of ML to astrophysical problems, and the recent adaptation of Transformer-based models for astronomical data, exploratory hybrid-model approaches to astrophysics can be justified.

4.1 Convolutional Neural Networks for Morphology Classification

One of the pioneering applications of ML with astronomy was the use of convolutional neural networks (CNNs) to classify galaxy morphology. Dieleman et al. [12] developed a rotationally invariant CNN with the intention to categorise galaxies based on visual morphology, achieving a performance comparable to human classification. Their work demonstrates the nature of deep learning architectures that extract complex spatial features from astronomical imaging data, developing key steps towards automated sky survey analysis.

4.2 Deep Learning for Photometric Redshift Estimation

Photometric redshift estimation is a crucial task for large-scale cosmological surveys where the introduction of ML approaches also provided benefit. D'Isanto and Polsterer [11] showed that deep neural networks could predict redshifts from photometric data with higher precision than traditional template-fitting methods. The model used provided a scalable and data-driven alternative, capable of handling a high volume of sources produced from surveys such as the Sloan Digital Sky Survey (SDSS)[10].

4.3 The Emergence of Transformer Models

While CNNs have proven highly effective for matrix based inputs like images or spectral data, the development of the Transformer architecture by Vaswani et al. [14] revolutionised machine learning by introducing self-attention mechanisms. Transformers model global dependencies across input features without relying on fixed-size convolutional kernels, making them exceptionally powerful for highly contextualised data. Their ability to cross reference all features with an input makes them especially useful in predicting outcomes in contexts when mapping between observational features and target variables is non-trivial, highly non-linear, or poorly understood [14, 15].

4.4 The Astronomy Transformer: A New Frontier

Recently, Luo et al. [15] introduced a generalised Transformer model tailored for analysing observational astronomical data. The Astronomy Transformer demonstrates high functionality across multiple tasks, including source classification, photometric redshift estimation, and anomaly detection. Its architecture is capable of handling heterogeneous inputs while capturing complex feature relationships, offering significant advantages over both traditional deep learning methods.

Success of Transformer-based models in astronomy highlight their potential as a modeling for astrophysical research. In particular, their ability to capture non-linear, global relationships between observational features suggests they are well-suited to enhancing the modeling of galaxy dynamics via dark matter distributions, forming the motivation for model integration into classical analytic frameworks.

5 Case Study: Hybrid Transformer Model for Galaxy Rotation Curves

To investigate the applications of Transformers within a specific field of astrophysics, the integration of transformer models with classical halo models was conducted. Requirements for this investigation include a feature representative dataset and custom built model.

5.1 Dataset Construction

To develop a dataset suitable for training, galaxy rotation curves from multiple sources, including SPARC[9]-derived measurements, were compiled to make up features and targets. Each galaxy in the dataset is represented with tabulated values for galactic radius R (in kpc), gas contribution V_{gas}, stellar disc contribution V_{disc}, bulge contribution V_{bulge}, and the HI linewidth measurement W_{50}.

Galactic radius R references the distance from the centre of the galaxy at which the features are measured. The gas contribution V_{gas} describes the rotational velocity induced by the Newtonian gravitational pull of the interstellar gas, primarily consisting of hydrogen. Stellar disc contributions V_{disc} account for the rotation velocity imposed by the mass of stars distributed within the galaxy's disc. The bulge contribution V_{bulge} corresponds to the gravitational influence of the central dense stellar bulge, which is typically significant at small radii. Finally, HI linewidth W_{50} captures the width of the 21-cm hydrogen emission line at 50% of its peak value, serving as a proxy of dynamic mass and rotational speeds of a galaxy.

The predictive target variable of the model is the observed rotational velocity V_{obs} as a function of radius. Prior to input into the model, galaxy features are normalised through the subtraction of the mean and division by the standard deviation of each feature across the dataset, ensuring zero mean and unit variance.

The dataset was split into training (80%), validation (15%), and test (5%) subsets, maintaining a balanced distribution of galaxy types across splits.

5.2 Hybrid Modeling Approach

Rather than directly predicting rotational velocities at each radius, the model uses a hybrid approach, learning to predict the parameters of two classical dark matter halo models — the Navarro-Frenk-White (NFW) profile [2] and the Einasto profile [4].

Specifically, the Transformer model outputs five parameters:

- NFW parameters: ρ_0 (central density) and r_s (scale radius),
- Einasto parameters: ρ_0 (central density), r_e (effective radius), and α (shape parameter).

Using these predicted parameters, the model computes rotational velocities $V(r)$ from analytical NFW and Einasto mass profiles. The final predicted velocity is taken as the product of a learnable parameter, and the mean of the two computed profiles, aiming to correct for systematic offsets. This hybrid strategy preserves the interpretability of classical models while allowing data-driven optimisation of their parameters with reference to galaxy properties.

5.3 Transformer-Based Model Architecture

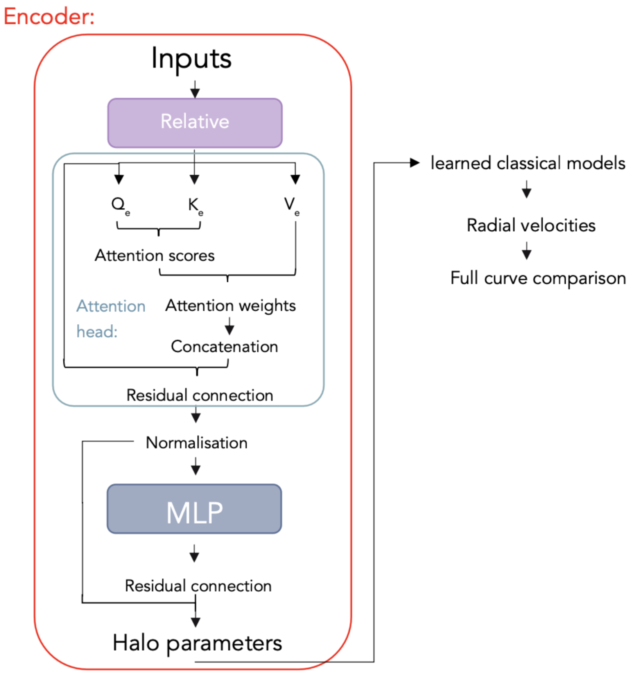

Figure 1: Full Transformer used for rotation curve modeling based on the architecture showcased by Vaswani et al. [14]

The fundamental driver of the model shown in Figure 1 is a Transformer encoder, originally introduced by Vaswani et al. [14], adapted for an astrophysical context following the generalised methodology outlined by Luo et al. [15].

The model architecture consists of:

- A linear embedding layer projecting the five input features into a higher-dimensional latent space.
- A Transformer encoder comprising two stacked layers, each with 4 attention heads and a feedforward dimension of 128.
- A final fully-connected layer that projects the latent representations into the five halo model parameters.

Self-attention mechanisms allow the model to dynamically weight the importance of different features at different spatial positions, capturing non-local dependencies across a galaxy's rotation curve. Their ability to cross-reference all features in parallel is further benefited through the use of multi-head mechanisms where one head may analyse the relationship between gas and bulge contribution while another compares Stellar disc contributions to HI linewidth [14, 15].

A notable feature of this model is that it operates on an entire galaxy at once, treating the sequence of radial measurements in full representation with each other, rather than as a temporal sequence. This enables the model to learn not just local trends, but the global structure of rotation curves across different radii.

5.4 Training Procedure

Model training was conducted with the use of PyTorch Lightning [16] to manage optimisation, checkpointing, and performance logging. The loss function, defined as the mean squared error between predicted and observed rotational velocities, with a $\log(1+x)$ transformation is applied with aims to reduce the influence of extreme values.

An Adam optimiser was employed with an initial learning rate of 10^{-3}, as training was performed over 200 epochs. The model checkpoint with the minimum validation loss was used for final evaluation.

6 Results and Discussion

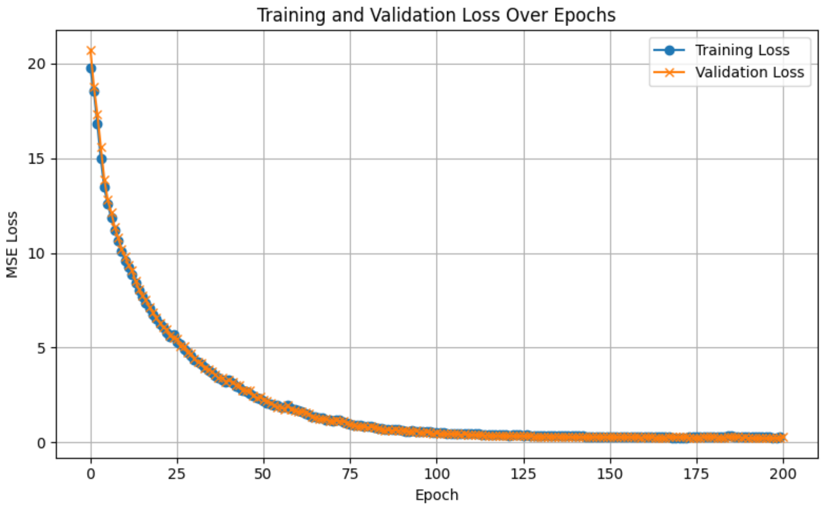

Figure 2: Training and validation loss plotted against 200 epochs, showing stable convergence and low final loss values = 0.08.

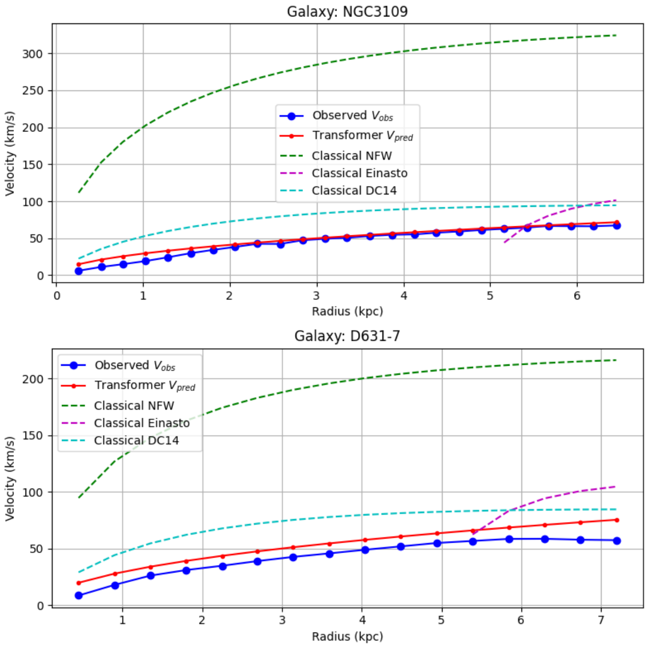

Figure 3: Comparison of observed and predicted rotation curves for unseen well-modeled galaxies NGC3109 and D631-7. The Transformer model tracks the observational data, closer than pure analytical models, across regions of radii between approximately 0.3 - 7 kpc.

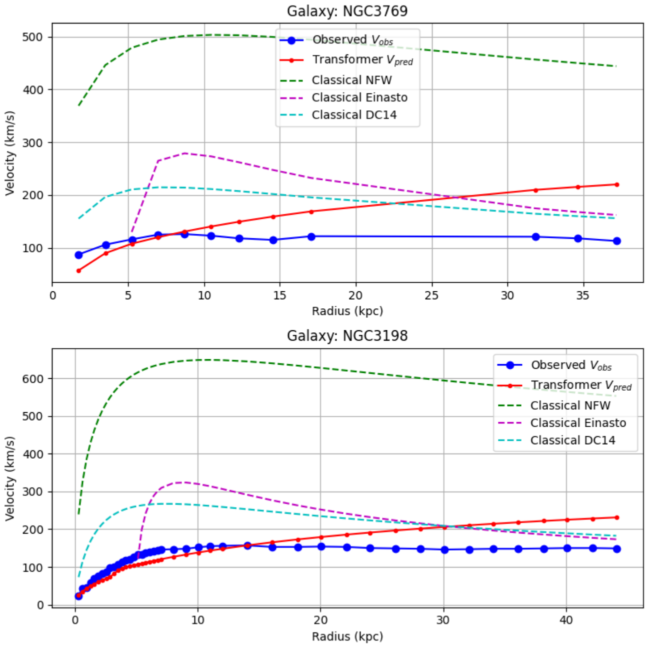

Figure 4: Comparison of observed and predicted rotation curves for unseen galaxies NGC3769 and NGC3198, showing increased deviation between model predictions and observations at larger radii greater than 20 kpc. These discrepancies reflect the challenges associated with larger, more complex galaxies.

The training performance of the model, shown in Figure 2, demonstrates a smooth and consistent convergence over 200 epochs, with both training and validation losses decreasing steadily before stabalising at a low loss. The absence of significant overfitting - shown via stabalised validation loss - suggests that the Transformer model is capable of generalising to unseen galaxies, as benefited from the self-attention mechanisms that allow complex, non-local feature dependencies to be learned, findings which are supported by A. Vaswani et al. (2017) and W. Luo, et al (2023) [14, 15].

For the model prediction cases such as NGC3109 and D631-7 (Figure 3), the Transformer-based model closely tracks the observed rotation curves across all low-radius galaxies (approximately 0 - 7 kpc), outperforming classical models which tend to systematically overestimate velocities. The physical properties of these galaxies, provided by F. Lelli, et al (2016)[9], are characterised with relatively low stellar masses and moderate HI linewidths. K. A. Oman et al (2015)[7] associate such properties with smoother, more coherent rotation curves, drawing parallels to the curves shown in Figure 3 and further validating the hybrid modeling methodology.

Conversely, model performance deteriorates at large radii galaxies with certain high-masses, such as galaxy NGC3769 and NGC3198 (Figure 4). In these cases, classical models typically overestimate velocities at radii of 10 - 20 kpc, but converge with observed velocities near the 30 - 40 kpc region. Notably, the Transformer predictions retain valid representation at low radii but show systematic deviations beyond ~ 20 kpc, suggesting that the training data may insufficiently capture the dynamic complexities present in massive, baryon-dominated systems. These failures may arise due to factors such as:

- Complex baryonic feedback processes (e.g., radiative gas cooling or strong stellar winds) not represented in input freatures, similarly to the works of W. J. G. de Blok (2010). [6],
- Asymmetries, or non-circular motions affecting the observed velocities, V_{obs} [8],
- Larger variance in stellar mass-to-light ratios among more massive galaxies, introducing degeneracies unaccounted for with feature normalisation.

Overall, the Transformer model exhibits a strong potential towards the improvement of galaxy rotation curve modeling, through its accurate representation of dark matter halos at low radii, compared to classical fixed-form profiles at central regions of a galaxy. Their ability to cross-reference multiple observational features in parallel, for the dynamic adaption of halo parameters highlights a significant advance toward addressing the classical modeling failures previously highlighted. Future improvements to a more generalist model could explicitly involve the incorporation of asymmetry parameters, non-circular motion indicators, or descriptions of the environment into model inputs, as followed by recent trends in hybrid astrophysical modeling introduced by W. Luo, et al (2023)[15]. Moreover, a hard cutoff at approximately 20 kps aiming to switch from the transformer based approach to classical methods could be introduced as part of a finalised deployed pipeline.

References

[1] V. C. Rubin and W. K. Ford, “Rotation of the Andromeda Nebula from a Spectroscopic Survey of Emission Regions,” *Astrophysical Journal*, vol. 159, p. 379, 1970. DOI: 10.1086/150317.

[2] J. F. Navarro, C. S. Frenk, and S. D. M. White, “The Structure of Cold Dark Matter Halos,” *Astrophysical Journal*, vol. 462, p. 563, 1996. DOI: 10.1086/177173.

[3] J. Einasto, “On the Construction of a Composite Model for the Galaxy and on the Determination of the System of Galactic Parameters,” *Trudy Astrofizicheskogo Instituta Alma-Ata*, vol. 5, p. 87, 1965.

[4] D. Merritt, J. F. Navarro, A. Ludlow, and A. Jenkins, “A Universal Density Profile for Dark and Luminous Matter?,” *Astrophysical Journal Letters*, vol. 624, pp. L85–L88, 2005. DOI: 10.1086/430636.

[5] A. Di Cintio, C. B. Brook, A. V. Macciò, G. S. Stinson, A. A. Dutton, and J. Wadsley, “The dependence of dark matter profiles on the stellar-to-halo mass ratio: a prediction for cusps versus cores,” *Monthly Notices of the Royal Astronomical Society*, vol. 437, pp. 415–423, 2014. DOI: 10.1093/mnras/stt1930.

[6] W. J. G. de Blok, “The Core-Cusp Problem,” *Advances in Astronomy*, vol. 2010, p. 789293, 2010. DOI: 10.1155/2010/789293.

[7] K. A. Oman et al., “The unexpected diversity of dwarf galaxy rotation curves,” *Monthly Notices of the Royal Astronomical Society*, vol. 452, no. 4, pp. 3650–3665, 2015. DOI: 10.1093/mnras/stv1504.

[8] S. Courteau et al., “Galaxy Masses,” *Reviews of Modern Physics*, vol. 86, pp. 47–119, 2014. DOI: 10.1103/RevModPhys.86.47.

[9] F. Lelli, S. S. McGaugh, and J. M. Schombert, “SPARC: Mass Models for 175 Disk Galaxies with Spitzer Photometry and Accurate Rotation Curves,” *Astronomy & Astrophysics*, vol. 615, p. A3, 2018. DOI: 10.1051/0004-6361/201630186.

[10] D. G. York et al., “The Sloan Digital Sky Survey: Technical Summary,” *Astronomical Journal*, vol. 120, p. 1579, 2000. DOI: 10.1086/301513.

[11] A. D’Isanto and K. Polsterer, “Photometric redshift estimation via deep learning,” *Astronomy & Astrophysics*, vol. 609, p. A111, 2018. DOI: 10.1051/0004-6361/201731189.

[12] S. Dieleman, K. W. Willett, and J. Dambre, “Rotation-invariant convolutional neural networks for galaxy morphology prediction,” *Monthly Notices of the Royal Astronomical Society*, vol. 450, no. 2, pp. 1441–1459, 2015. DOI: 10.1093/mnras/stv632.

[13] D. George and E. A. Huerta, “Deep Neural Networks to enable real-time multimessenger astrophysics,” *Physical Review D*, vol. 97, no. 4, p. 044039, 2018. DOI: 10.1103/PhysRevD.97.044039.

[14] A. Vaswani et al., “Attention is All You Need,” in *Advances in Neural Information Processing Systems (NeurIPS)*, vol. 30, 2017.

[15] W. Luo, M. Wu, W. Xu, X. Liang, and X. Wang, “Astronomy Transformer: A Generalized Transformer Model for Astronomical Observation Data,” *arXiv preprint arXiv:2310.12069*, 2023. Available: arXiv:2310.12069.

[16] W. Falcon, “PyTorch Lightning,” GitHub repository, 2019. Available: https://github.com/Lightning-AI/lightning.

[17] K. C. Freeman, “On the disks of spiral and S0 galaxies,” *The Astrophysical Journal*, vol. 160, pp. 811–830, 1970. DOI: 10.1086/150474.