

RELATIONAL REASONING

The background is a solid blue color with a white 'X' shape formed by two diagonal lines. Scattered across the background are various 3D geometric shapes in red, blue, yellow, and purple. These include spheres, cubes, and cones. Some shapes are clustered together, while others are isolated. The lighting creates soft shadows for the 3D objects.

*CS337: Artificial Intelligence
and Machine Learning*

KARAN GODARA (210050082)

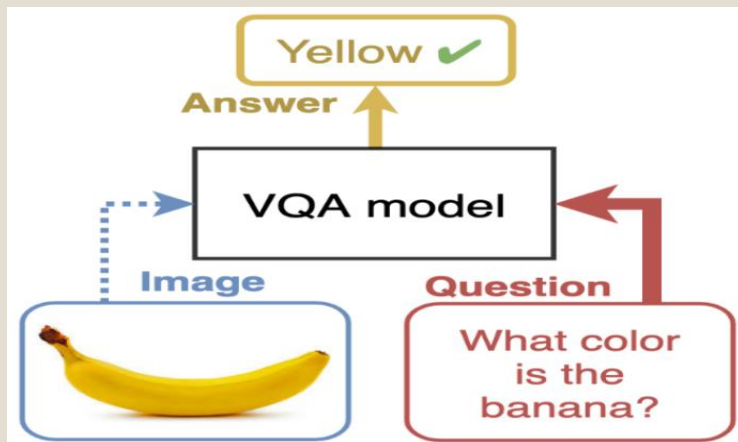
ISHA ARORA (210050070)

SHANTANU WELLING (210010076)

DHANANJAY RAMAN (210050044)

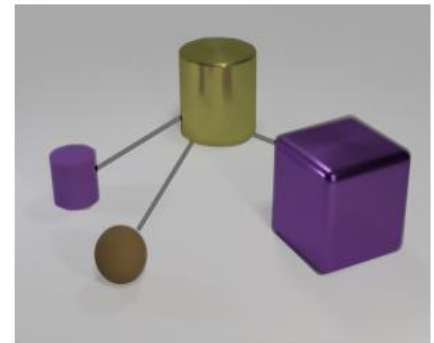
Project Overview

- **Visual Question Answering (VQA)** : ML task to understand natural language questions about an image and provide accurate answers.
- **Relation Network (RN) Model** : A general solution to flexible *relational reasoning* in neural networks



Relational question:

Are there any rubber things that have the same size as the yellow metallic cylinder?



Project Overview

- ***Our Experiments :***

- Conducted experiments on both CLEVR and Sort-of-CLEVR datasets
- Explored relation networks, focusing on their flexibility through the incorporation of various descriptions of 'objects'
- Enhanced the model's performance on Sort-of-CLEVR dataset

Why previous approaches fail?

- Perform well on non-relational questions but perform below par in the relational question domain.
- Symbolic approach to AI is inherently relational and not robust to task and input variations.
- Relational question domain demands mechanisms to compute relations between a set of objects.
- **CNN** – capacity to reason about spatial, translation invariant properties
- **RNN** - capacity to reason about sequential dependencies
- Now let's explore RN architecture which has the capacity to compute relations baked into them.

Relation Network

- **Design philosophy** - constrain the functional form of a neural network so that it captures the core common properties of relational reasoning

$$\text{RN}(\mathcal{O}) = f_{\phi} \left(\sum_{i,j} g_{\theta}(o_i, o_j) \right)$$

- input is a set of “objects” $\mathcal{O} = \{\mathbf{o}_1, \mathbf{o}_2, \dots, \mathbf{o}_n\}$, $\mathbf{o}_i \in \mathbb{R}^m$ is the i^{th} object, and \mathbf{f}_{ϕ} and \mathbf{g}_{θ} are functions with parameters ϕ and θ , respectively.
- In our case, \mathbf{f}_{ϕ} and \mathbf{g}_{θ} are MLPs.

Strengths of RNs

1

Learn to infer relations: Considers the potential relations between all object pairs

2

Data efficient : Use a single function g_θ to compute each relation – greater generalisation

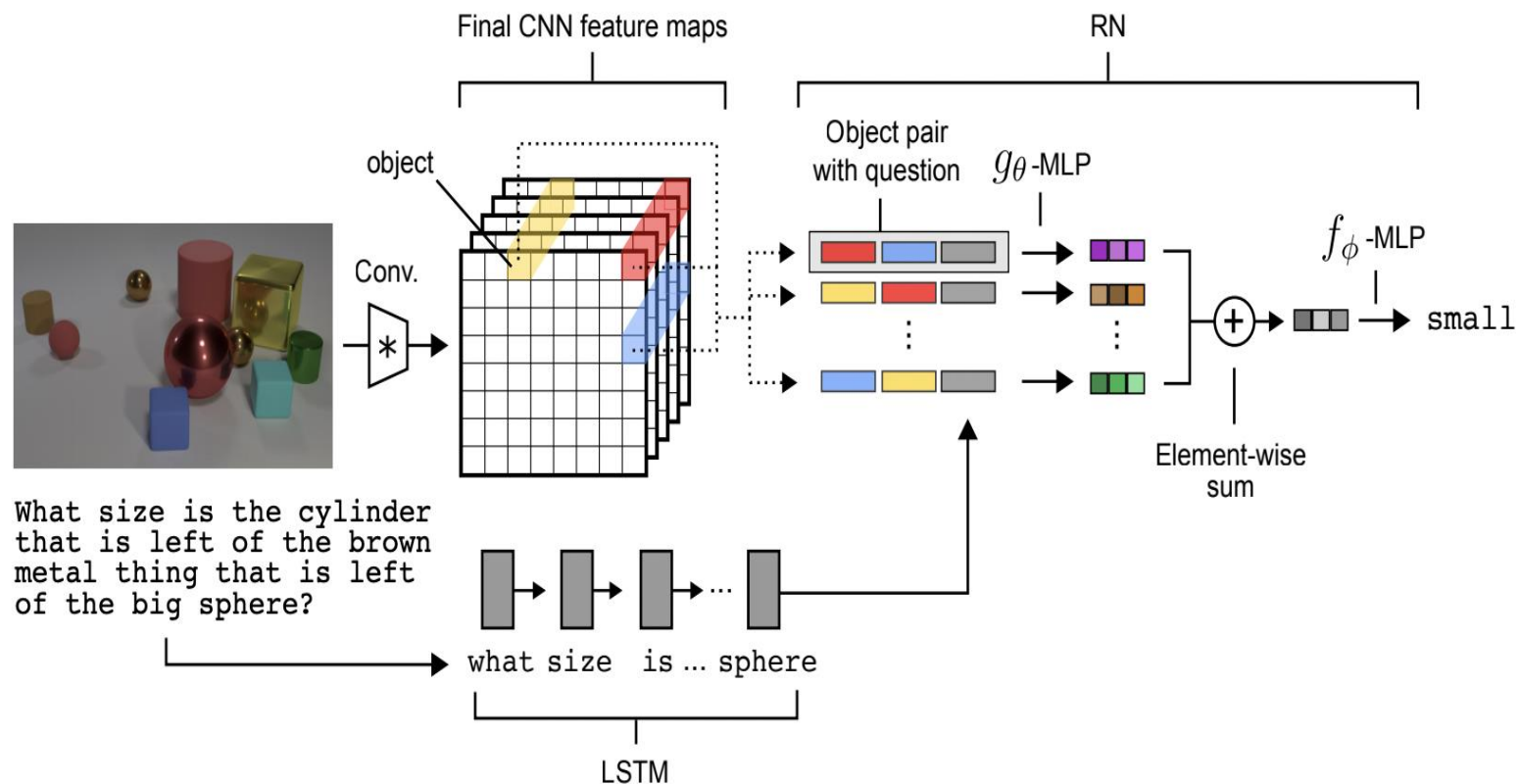
3

Operate on a set of objects : Invariant to the order of objects in the input

4

Flexible : Do not explicitly operate on images or natural language

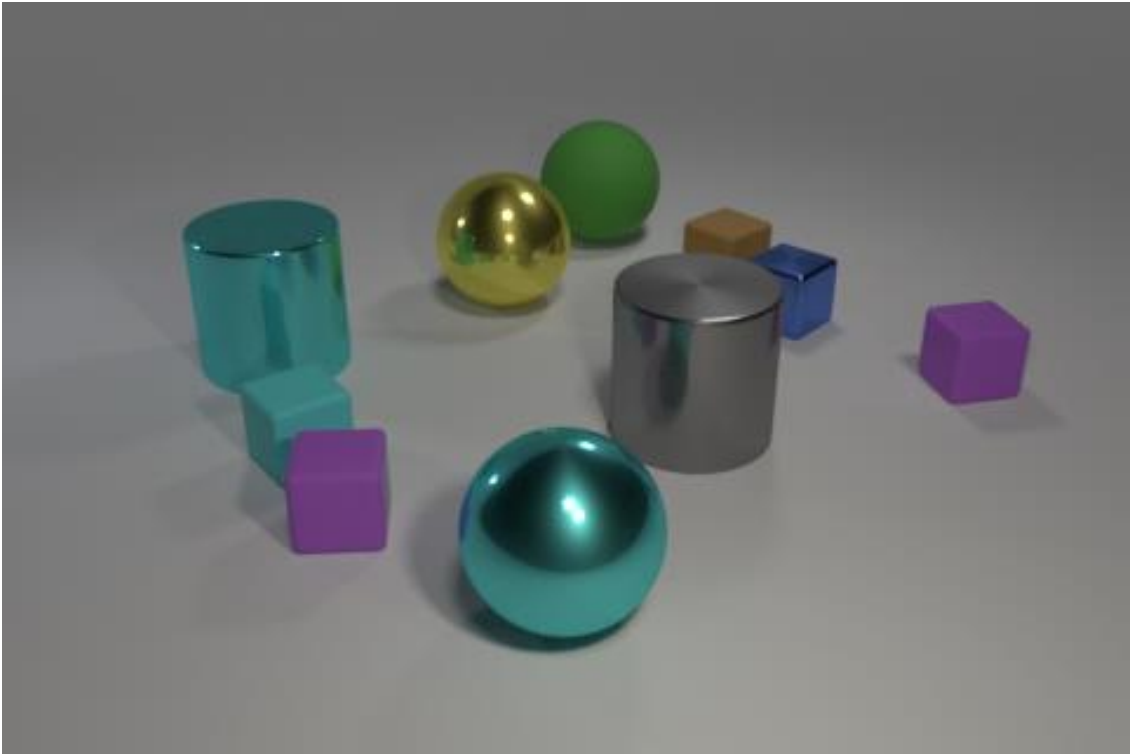
Relation Network Architecture for Images



Datasets Employed

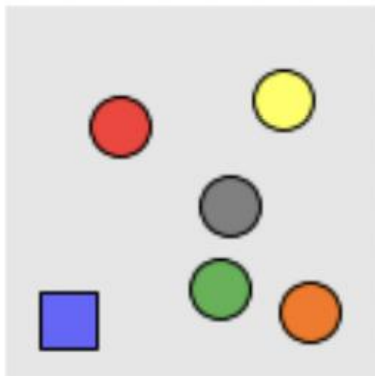
***CLEVR (Compositional
Language and Elementary
Visual Reasoning) :***

VQA dataset of images of
3D objects on which state-
of-the-art approaches have
struggled due to the
demand for rich relational
reasoning – spatial and
otherwise



Datasets Employed

- **Sort-of-CLEVR** : Dataset similar to CLEVR but consists of only 2D color shapes that separates relational and non-relational questions



Non-relational question

Q: What is the shape of the red object?

A: circle

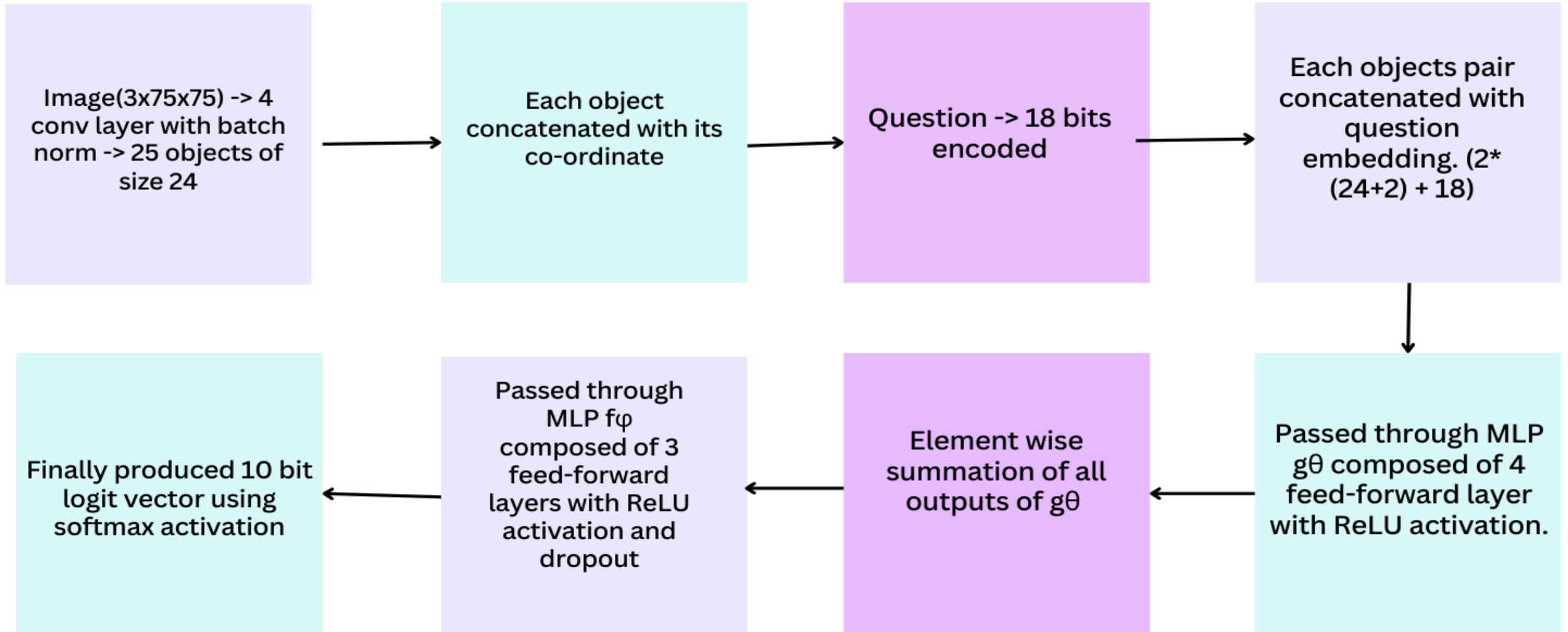
Relational question

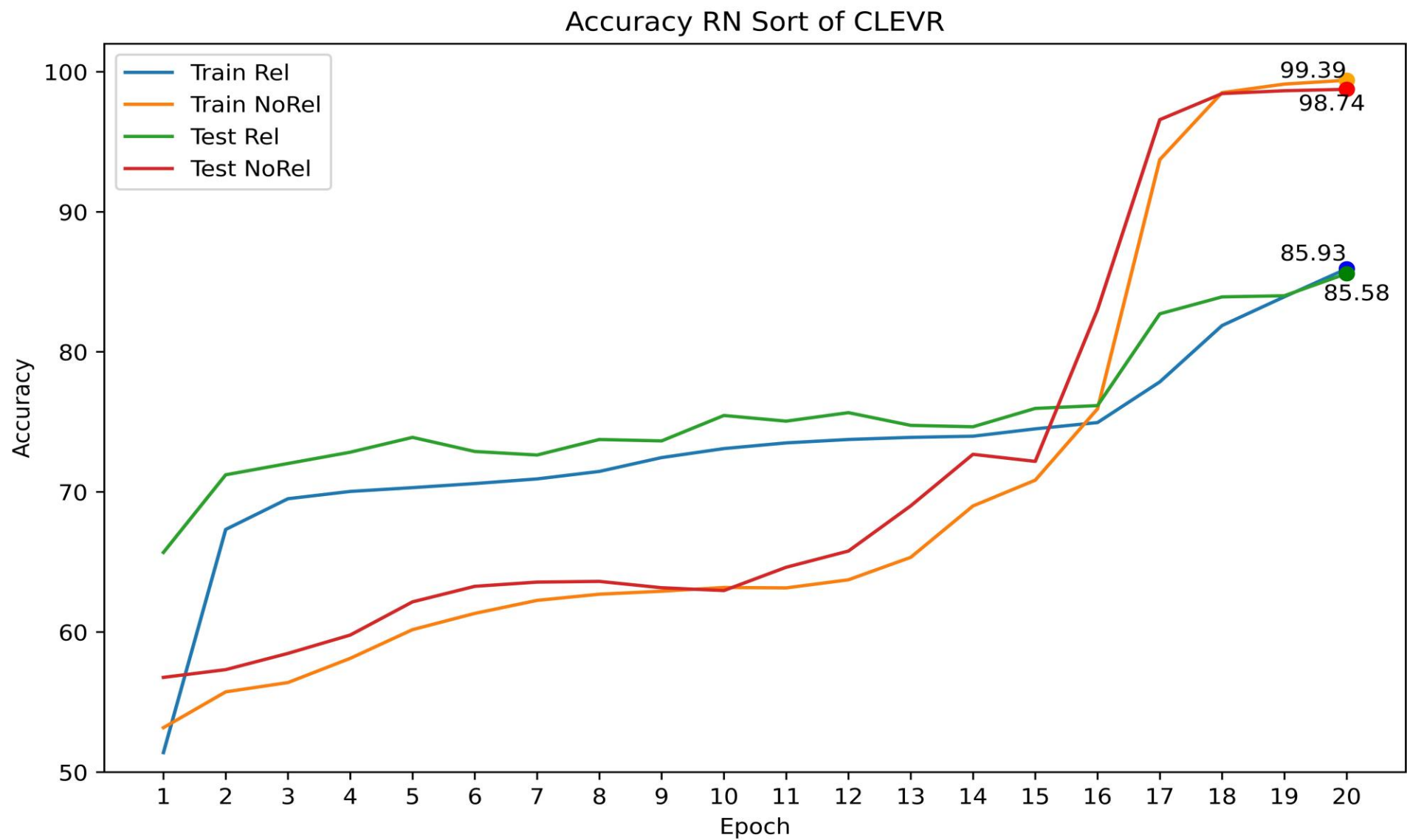
Q: How many objects have the shape of the blue object?

A: 1

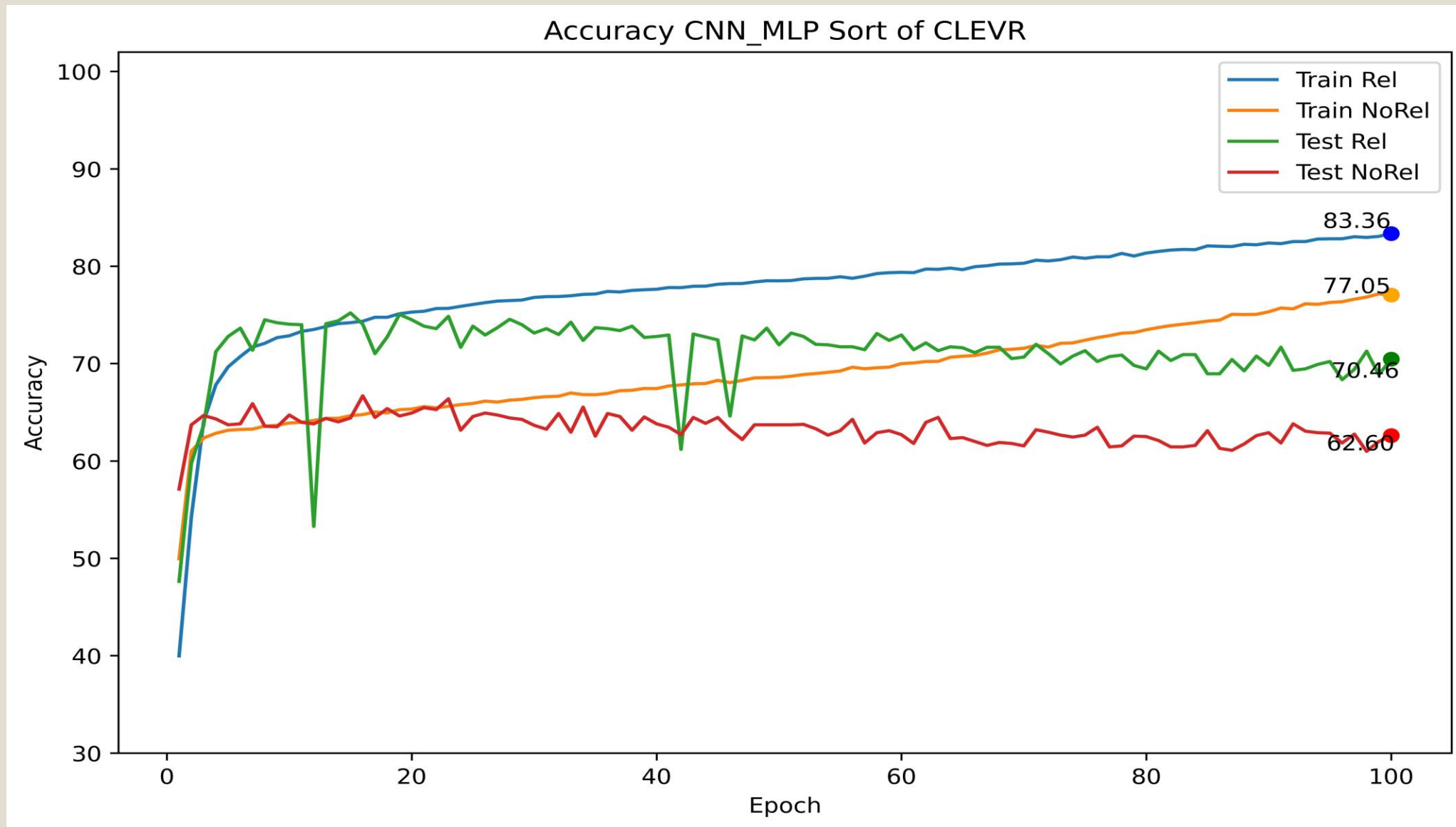
OUR EXPERIMENTS

Sort-of-CLEVR with RN





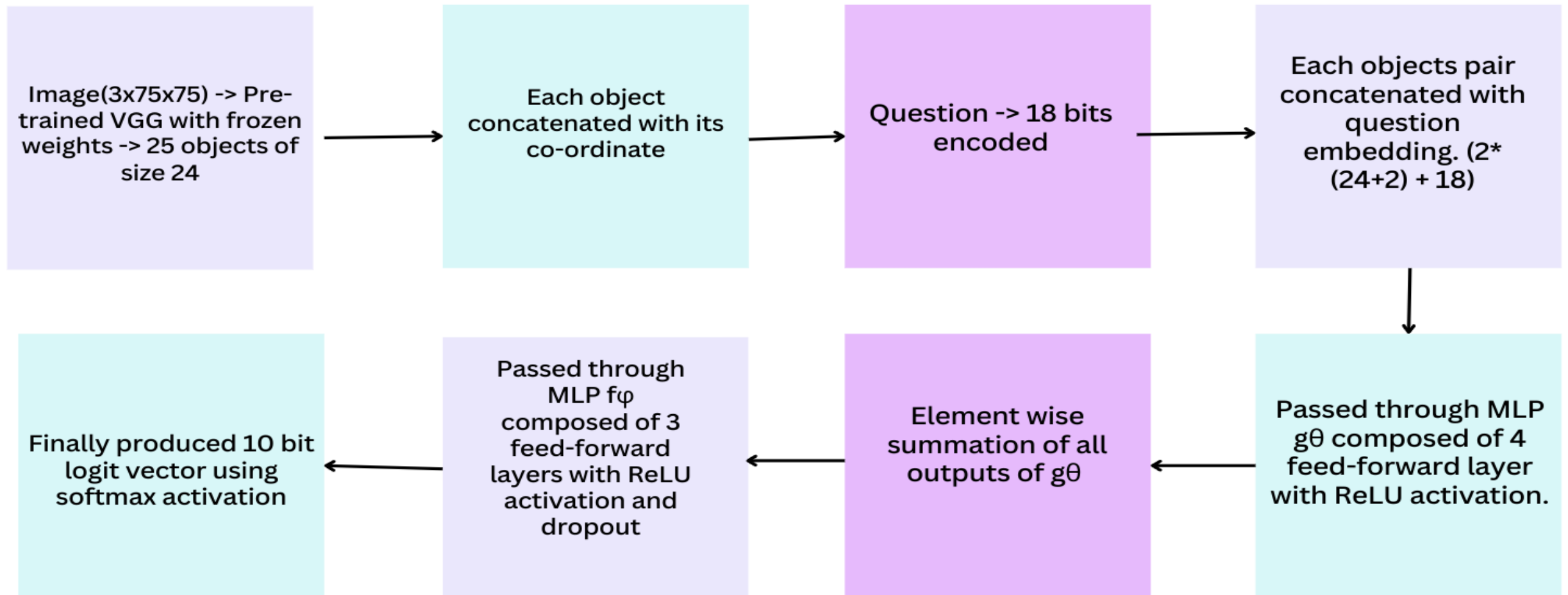
RN on Sort-of-CLEVR data

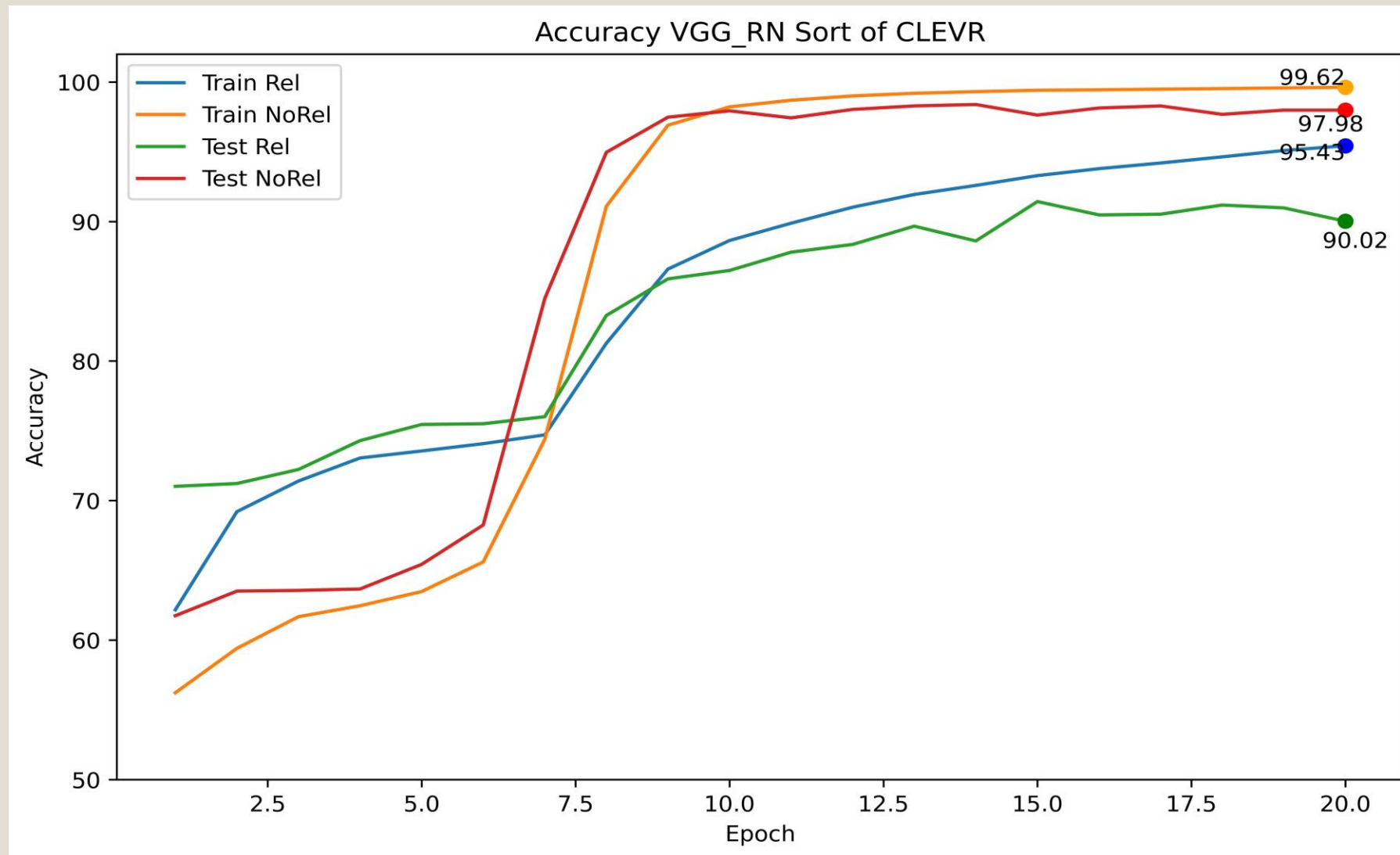


Relation Network Model outperforms the CNN MLP model

Changing object generator CNN

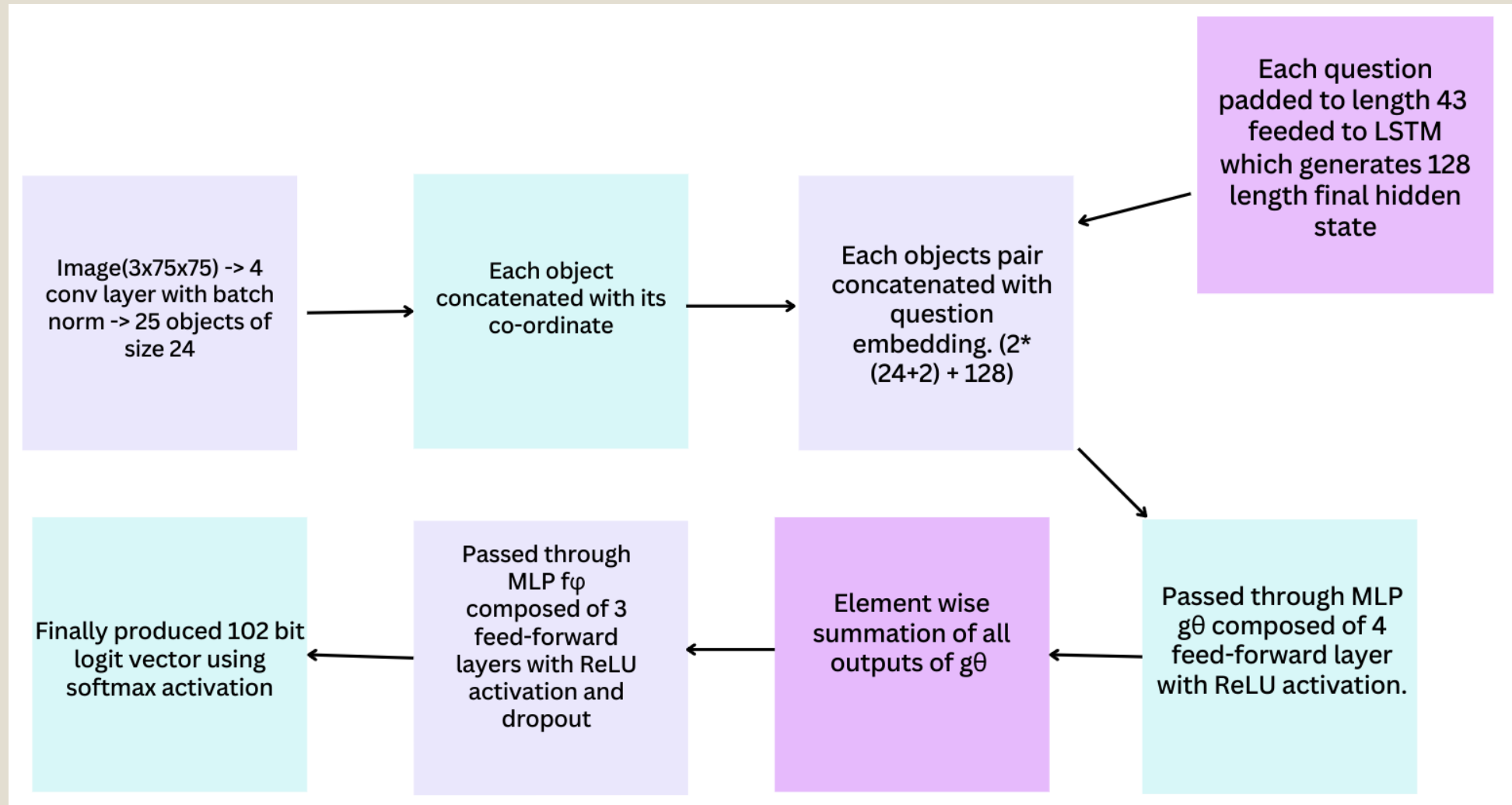
Sort-of CLEVR dataset with pre-trained

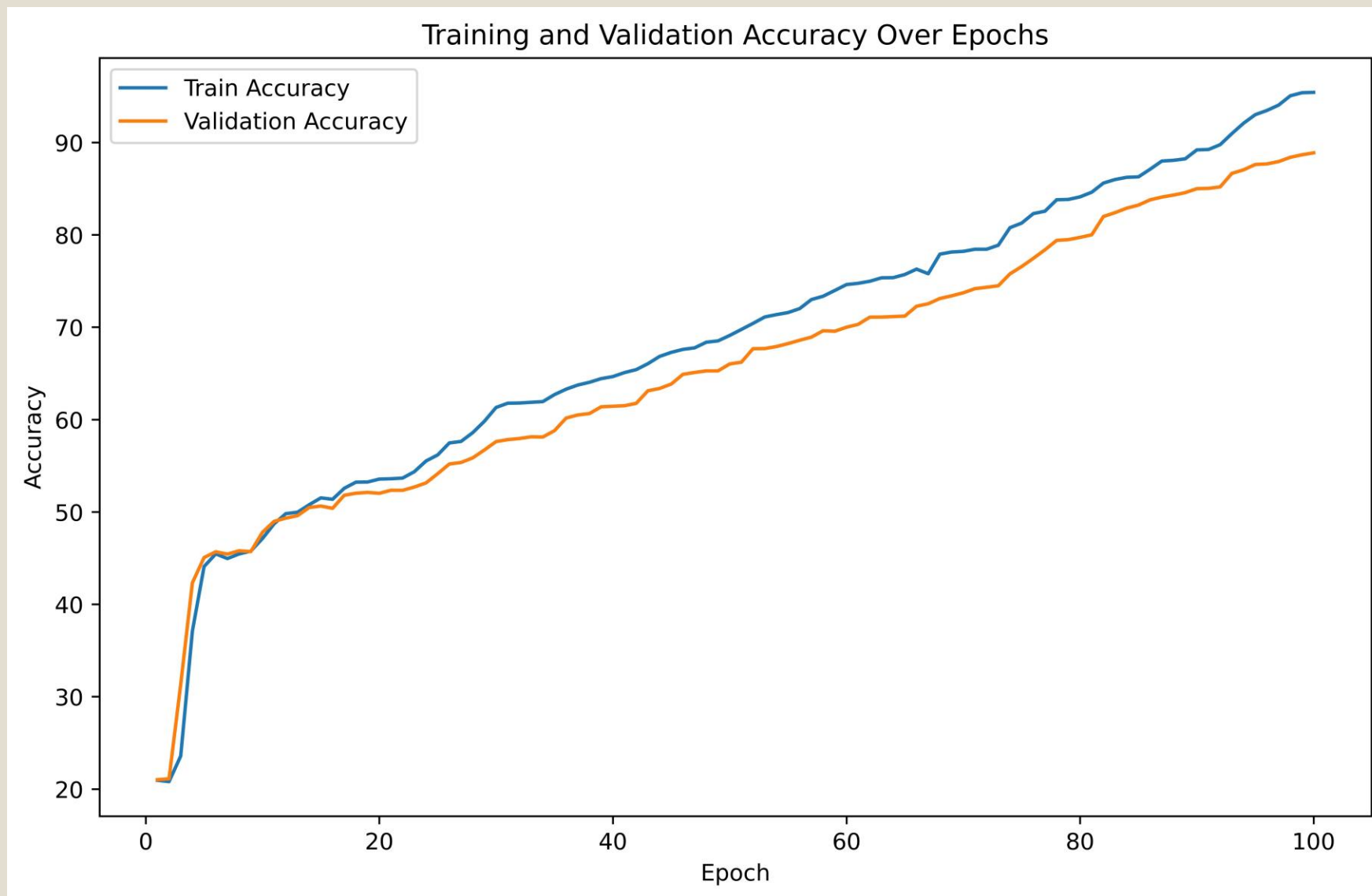




Improvement in performance by using pre-trained VGG based RN on Sort-of-CLEVR Data

CLEVR Dataset





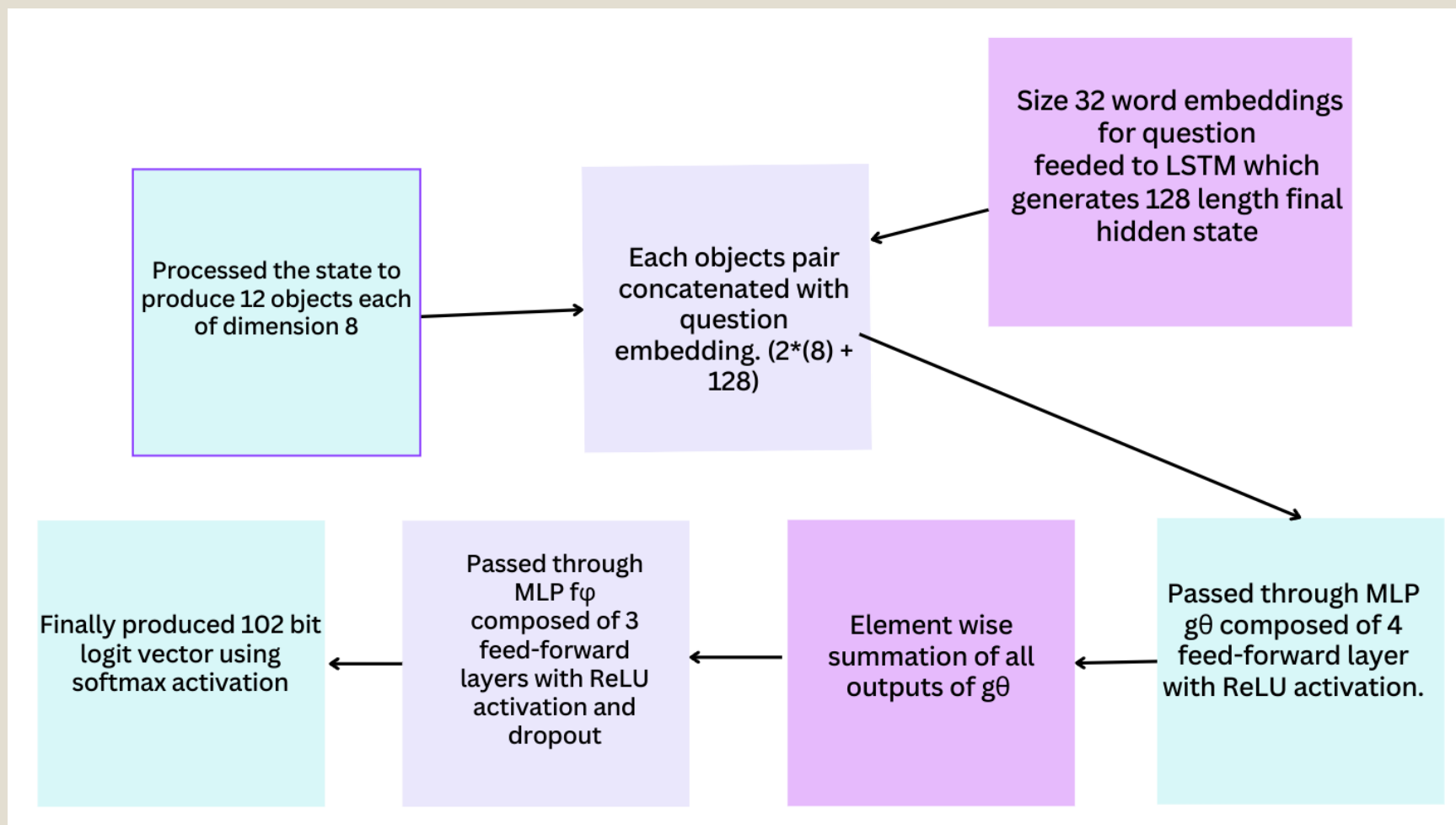
RN on CLEVR Dataset (Pixel Representation)

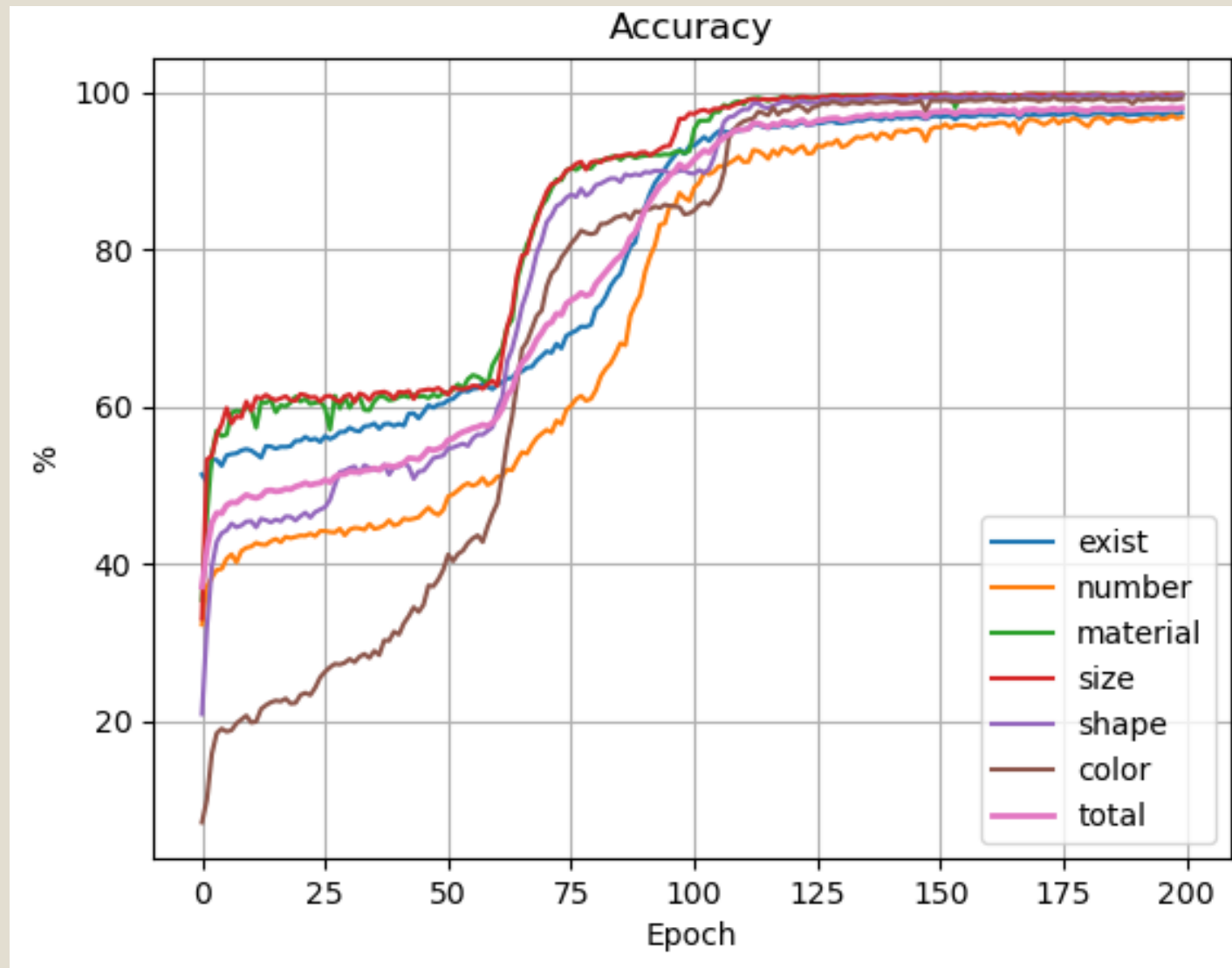
Versatility with regards to form of input

CLEVR Dataset with State Description

- Images explicitly represented by state description matrices containing factored object descriptions
- Each row in the matrix contains the features of a single object
 - 3D coordinates (x, y, z)
 - color (r, g, b)
 - shape (cube, cylinder, etc.)
 - material (rubber, metal, etc.)
 - size (small, large, etc.)

CLEVR Dataset with State Description





RN model on CLEVR Data with state description as input

Further Improvements



Inclusion of ternary relations



Employ deep CNNs with attention to improve object representation



RN definition can be adjusted to consider only some object pairs. RNs can take as input a list of only those pairs that should be considered, made available by some upstream mechanism

References

- <https://github.com/kimhc6028/relational-networks>
- <https://github.com/mesnico/RelationNetworks-CLEVR/tree/master>
- <https://arxiv.org/pdf/1706.01427.pdf>

Contributions

Each individual experiment's codebase was developed majorly by one individual with the assistance of others. Overall, everybody had the same contribution.

The background of the slide is a complex, abstract geometric pattern. It consists of numerous thin, colored sticks (blue, green, yellow, and red) that are interconnected at various points using small, yellow, three-pronged connectors. These sticks and connectors form a dense, overlapping network of geometric shapes, primarily triangles and quadrilaterals, creating a sense of depth and complexity. The overall color palette is muted, with the sticks being a darker shade of their respective colors.

Thankyou

KARAN GODARA (210050082)

ISHA ARORA (210050070)

SHANTANU WELLING (210010076)

DHANANJAY RAMAN (210050044)