

1 ————— To Reviewers R1, R2, R3, R4 —————

2 We thank all the reviewers for their insightful comments. In particular, we sincerely appreciate that reviewers (R1,R3,R4) acknowledge our proposed approach to be well-motivated, novel, and interesting. To improve our paper, we took the reviewers’ advice to clarify and add details in our explanation for better understanding (R1, R2, R4), provide new insights to regarding the amount of data in visual reasoning problems (R3), and perform exhaustive experiments (R2). Finally, as suggested by (R1, R2), we provide experimental results on the PGM and Balanced-Raven dataset (see Table A,B).

9 (R1,R2,R4) **Details in Eq.(2,3,4) in Section 3.2.** We apologize for the oversimplified explanation. As mentioned in the manuscript, an analogy is learned by minimizing the difference between the (modelled) distributions over the objectives (e.g. Inference probability for task \mathcal{T} , $Pr(\mathcal{T}|S)$, hereafter $p(S)$) over transformed variables (S , embedding vector by CNN). Since the objectives have no ground truths, we use the estimated distribution from the support context as the reference. In short, we enforce the distributions over random variables which embed analogical relation to be similar. @R1 The distribution is modelled by a multivariate transformation (e.g. encoder). For empirical probability, a softmax function is followed by output of the transformation. We will include the terminology and notation in detail in the revision. @R2 A summation of these multidimensional BCE output returns a scalar value. $p(S)$ and $q(Q_i)$ indicate these empirical probabilities. @R4 Notations ‘ p ’ and ‘ q ’ simply follow the convention for BCE equation. We will include further explanations in the revision.

19 (R1, R2) **PGM.** As suggested, we experimented on PGM over a baseline (WRen) (Table A above), and the results show that our method outperforms it. We will include the results with more details in the revision.

21 (R1) **Connect the proposed method to analogy.** Our method include two analogies: (a) within the context panels and (b) between the context data (support/query). (a) is explained with relational sets $\{\mathbf{x}_i : \mathbf{x}_j :: \mathbf{x}_r : \mathbf{x}_t\}$, where \mathbf{x} denotes relational subsets (\mathbf{x} : combination of panels) of the context data X . (b) is explained with X and \tilde{X} (perturbed X) and their embedding vectors S and Q . Hence, we rephrase Eq.(1) with an analogy set $A = \{S :: Q_i\}_{i=1}^8$. Analogy losses (maximum likelihood) assist to learn their analogies between modelled distributions. (R1) **Clarification.** ¶The four-word proportional analogy is merely a simple example and does not explain all the cases, as described in the manuscript. ¶L85: as noted in the paper, as the affine-transform based rule is a simple example but not a generalized one, we will eliminate it to avoid misleading. ¶Eq.(1): we will revise it to more direct expression ($\bigcup_i \{S :: Q_i\}$), which is also an analogy. ¶Kernel: we will replace this term with “similarity measure”. ¶Analogy losses: as long as S and Q are highly related, analogy learning using probability models will work regardless of their independent relations.

31 (R2) **Argument on novelty.** Our novelty is the proposal of meta-analogical learning for efficient reasoning over limited observations. VAE and NCE, which R2 pointed out, are auxiliary or partial components to realize our approach. Our proposed framework is in a general scope so that it can be easily combined with the existing learning framework (e.g. MAML) as shown in the supplementary material. (R2) **Balanced-RAVEN.** Thank you for informing us about this dataset. Our method outperforms CoPINet on this dataset (Table B).

36 (R2) **More related works.** Most related works are explained in the Preliminaries section in L90-109. We will add a discussion with Hu *et al.* paper in the revision. (R2) **Clarification.** ¶Encoder and an inference network: as explained in L157, the task inference network uses the encoder. We will eliminate L158 for better clarity. ¶‘:’ and ‘::’: even though it is explained in L33-35, we will add explanation (‘:’ - relation and ‘::’ - analogy) for clarity. ¶Grammar: we will proof-read it again. We will revise the notations as suggested.

43 (R3) **Criticisms and argument about data size.** Thank you for your insightful comments. 1) Although we were referring to the number of samples by the term “large”, we agree that each sample has limited information since it contains small number of pixels, and will tone down on the strong arguments in [L2-3] and [L29-30]. 2) We will clarify that our method leverages the data in a more efficient manner as it is guided by the underlying rules. 3) We agree and believe that the analogy is one such example of such high-level abstraction, as it captures the relational similarities between inputs regardless of their raw input values.

49 (R4) **NCE.** As mentioned in the manuscript, NCE is a common learning objective and thus we did not emphasize it that much. We followed that of CoPINet: $\mathcal{L}_{nce} = \log(f(h_w(X, C(y)) - b)) + \sum_{y_i \neq y} \log(1 - (f(h_w(X, C(y_i)) - b)))$ where b is a constant. We will further clarify this term in revision. (R4) **Meta-analogical?** For meta-learning of our model, we use the dummy-task labels using the predefined annotation (attributes and rules) as shown in the supplementary material (Table 2. in Section A.), and test on unseen types of attributes or rules at meta-test time. (R4) **Dimensionality.** This is given in the Section B of supplementary material. (R4) **Literatures.** The main idea of our work is not the use of contrastive learning for visual reasoning, but is the proposal of meta-analogical learning. Yet, we will include the suggested references in the revision. ¶Figure 3. and 4.: as suggested, we will simplify the figures in the final version.

Table A: Improved performances on PGM.

#sams	Base	+Anal	+Anal+Inf	+Anal+Var	+Anal+Gen
585	12.70	13.28	13.44	13.46	13.81
1171	13.37	13.91	13.90	13.49	13.59
4687	12.76	13.34	14.17	14.49	14.49

Table B: Improved performances on Balanced-RAVEN.

#sams	Base	+Anal	+Anal+Inf	+Anal+Var	+Anal+Gen
14	14.77	15.71	15.51	14.81	15.11
35	15.70	17.17	14.81	16.42	15.52
77	15.57	15.93	15.81	15.96	14.74
322	16.66	16.75	17.11	16.37	17.31