Numerosity Representation in InfoGAN: An Empirical Study

Andrea Zanetti[1,2](✉), Alberto Testolin[3], Marco Zorzi[3,4],
and Pawel Wawrzynski[1]

[1] Warsaw University of Technology, Warsaw, Poland
pawel.wawrzynski@pw.edu.pl
[2] Intel Technology Poland, Gdansk, Poland
andrea.zanetti@intel.com, a.zanetti@ii.pw.edu.pl
[3] Department of General Psychology and Padova Neuroscience Center,
University of Padova, Padua, Italy
{alberto.testolin,marco.zorzi}@unipd.it
[4] IRCCS San Camillo Hospital, Venice, Italy

Abstract. It has been shown that *"visual numerosity emerges as a statistical property of images in 'deep networks' that learn a hierarchical generative model of the sensory input"*, through unsupervised deep learning [1]. The original deep generative model was based on stochastic neurons and, more importantly, on input (image) reconstruction. Statistical analysis highlighted a correlation between the numerosity present in the input and the population activity of some neurons in the second hidden layer of the network, whereas population activity of neurons in the first hidden layer correlated with total area (i.e., number of pixels) of the objects in the image. Here we further investigate whether numerosity information can be isolated as a disentangled factor of variation of the visual input. We train in unsupervised and semi-supervised fashion a latent-space generative model that has been shown capable of disentangling relevant semantic features in a variety of complex datasets, and we test its generative performance under different conditions. We then propose an approach to the problem based on the assumption that, in order to let numerosity emerge as disentangled factor of variation, we need to cancel out the sources of variation at graphical level.

1 Introduction

There is general consensus that humans's ability to perceive numerosity in visual stimuli relies on two core neuro-cognitive systems [2]: the Approximate Number System (ANS) enables to roughly estimate numerosity when there are many items in the visual display, whereas a second system processes small numerosities (in the "subitizing" range, typically up to four items) and it is tied to tracking objects in time and space. In order to represent numerical quantity at a semantic level, a cognitive system would need to abstract it away from the many low-level (e.g., graphical) features present in the sensory input, thereby extracting numerosity as a common factor of variation between the images.

© Springer Nature Switzerland AG 2019
I. Rojas et al. (Eds.): IWANN 2019, LNCS 11507, pp. 49–60, 2019.
https://doi.org/10.1007/978-3-030-20518-8_5

Recent simulation work [1,5,6] has shown that deep learning models can reproduce human performance in numerosity discrimination tasks that tap the ANS, suggesting that our numerical abilities might emerge from domain general learning mechanisms [3]. In particular, Stoianov and Zorzi [1] argued that in their model numerosity was *"computed through the combination of local computations and a simple global image statistic (cumulative area), without explicit individuation and size normalization of visual objects"*. However, despite these initial findings it is still unknown whether deep networks could learn to encode numerosity as a single explicit dimension, which would allow to control the generative process in an efficient and interpretable way, rather than having numerosity encoded as distributed pattern of activation of some neurons that can be "decoded" via the use of a trained linear classifier, with no direct access and control over it.

In the present work we address this question by focusing on InfoGAN [4], a powerful deep learning model that has been shown capable of disentangling the most relevant factors of variations in many different datasets, ranging from handwritten digits to faces [16]. We study if and how this type of generative model could learn to map numerosity into one or possibly more latent variables. Although it has been proved that it is theoretically impossible for an arbitrary generative model to learn disentangled representations of the input data in unsupervised settings [7], this impossibility holds *a priori* only for models without any inductive biases suitable for the task at hand (that is, the set of solutions the unsupervised model is able to produce and their probability under the model). Inductive biases can be expressed in many ways (model architecture, training algorithm, initialization scheme, etc.). In our study, we explore the role of different biases in the InfoGAN by adding cost components and varying the model architecture, latent space dimensionality and other hyperparameters.

Our main contributions can be summarized as follows. Three different models are analyzed with the aim of investigating the emergence of single elements of the latent code that would represent numerosity; the models are based on the following assumptions:

- with the first InfoGAN model, we implicitly make the assumption that no particular strategies must be considered to abstract numerosity from other graphical features in the input data in an unsupervised fashion;
- with the second model, we move to a semi-supervised setting to overcome the challenges resulting from a completely unsupervised learning regimen;
- with the third model, we tackle the problem of mapping numerosity as a disentangled dimension in the latent space assuming that this might emerge if we cancel out the sources of statistical variation at the graphical level.

Overall, one of the models appears to have the greater potential for disentangling numerosity. Our experiments also confirm the difficulties indicated in [5]. The outline of the paper is as follows: Sect. 2 overviews related literature. Section 3 formulates the problem. Section 4 presents experimental results, which are discussed in Sect. 5. The last section concludes the paper and outlines further research.

2 Related Work

An important assumption in representation learning [7,9] is that real-world data (like images or videos) can be thought as generated by a generative process that has two phases: first a latent random variable is sampled from a (possibly multivariate) prior distribution $P(z)$, where z can be thought of as the "cause" of the semantic factor of variations in the input data, and then the real-word data x would be generated by sampling a conditional probability $P(x|z)$. A classic unsupervised learning task is to find the "best" representation of the data, meaning a representation that embodies as much information about the input data as possible, being at the same time constrained to meet some conditions that are application-specific. In the present work, we have taken the position for which "best" corresponds to "disentangled", meaning a data representation that attempts to disentangle the sources of variation underlying the data distribution such that the dimensions of the representations are statistically independent [9].

The InfoGAN [4] is a variation of Generative Adversarial Networks [10] that extends the basic adversarial setup with a regularization based on the maximization of the Mutual Information between the Generator output and part of the latent code fed into the Generator itself. The InfoGAN model considers a minmax game that starts from the fundamental GAN minmax game:

$$\min_G \max_D V(D,G) \tag{1}$$

$$V(D,G) = \mathbb{E}_{x \backsim P_{data}}[\log(D(x))] + \mathbb{E}_{(c,z) \backsim P(c,z)}[\log(1 - D(G(c,z)))] \tag{2}$$

and adds a regularizer that represents the Mutual Information between the generated output $G(c,z)$ and the coding part c of the noise fed into the Generator to obtain a modified minmax game:

$$I(c; G(c,z)) = H(c) - H(c|G(c,z)) = H(G(c,z)) - H(G(c,z)|c) \tag{3}$$

$$\min_G \max_D V_I(D,G) \tag{4}$$

$$V_I(D,G) = V(D,G) - \lambda I(c; G(c,z)) \tag{5}$$

Also the related CatGAN model [11] is of interest here, as the author introduced the extension for the Discriminator to classify the output of the Generator either in a number of classes known a priori, when it is possible to access labels for the dataset at hand, or simply using an "estimated" number of classes. When the labels are accessible, it is possible to add a cost component corresponding to the cross-entropy between the sought distribution and the one obtained at the classifier output. When the labels are not available, by assuming a specific number of classes (but ignoring the labels for each entry) it is possible to add a cost component maximizing a ratio between the entropy of class y assigned to $G(c,z)$ by the classifier (which is supposed to be high) and the entropy of the assigned label conditioned to $G(c,z)$, so $y|G(c,z)$, (which is supposed to be low, for the choice to be sure); this is basically the same idea used in the Inception Score [12]. It can also be shown [13] that maximizing the Inception Score corresponds to maximizing the Mutual Information between the input being classified, in this case $G(c,z)$ and the class y being outputted by the additional classifier.

3 Problem Formulation and Methods

3.1 Problem Formulation

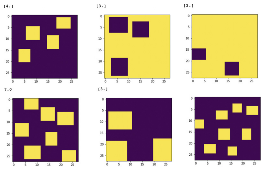

Fig. 1. Dataset samples.

The main problem considered here is to study **if** and **how** it is possible for the InfoGAN model to learn to represent numerosity as independent factor of variation in its latent space. To this aim, we extended the InfoGAN to make it develop the same disentangled representations seen on the MNIST dataset, but extracted from a dataset composed of images containing a different number of items [1] (see samples in Fig. 1). Though the visual structure of these images might look simpler than MNIST images, the most relevant *semantic* direction of variation is not strictly graphical but more abstract in nature, being indeed numerosity.

To test the hypothesis behind each model we consider, and assess the quality of the related learning process, we explicitly look for a "minimum degree" of disentanglement in the representation learned by each model, investigating what kind of latent space representation is induced by the learning process for each model, and we *visualize it* by changing one latent variable at a time and generating new data (images) with the trained Generator of the model under test. The visual inspection of the Generator output obtained in this way is a first qualitative indicator of whether any disentanglement has been reached in the process. Though this approach may seem fuzzy, there is no formal definition (yet) of disentanglement which is widely accepted, and there is no unified protocol to quantify it [7]; therefore, in this paper we accept the visual inspection as first qualitative evaluation of disentanglement of the latent space, as it has been done in the original InfoGAN paper [4].

3.2 Methods

We start out with the InfoGAN model as used in [4] (experiment 1 in that paper), but using a synthetic dataset obtained from the one used in [1] after applying the following basic transformations ad data augmentation techniques:

1. We reduce the numerosity range from 1 to 8, with most the experiments actually focusing on the range from 1 to 4. Though these intervals are somewhat limited, they still leave the possibility to build instructive parallels between the simulations and the ANS/Subitizing distinction.
2. We apply simple data augmentation procedures based on image reflection along orthogonal and diagonal axes.
3. We invert the background with the foreground, therefore doubling the total amount of images.

In so doing, we obtain a full dataset, for numerosity 1 to 8, composed of 128.000 images, half of in black over white, and half in white over black. In particular, the third transformation is motivated by the fact that numerosity should be minimally linked to any graphical representation, being a concept connected to “areas of coherence or correlation” in the perceptual input domain, whatever this could be (visual, audio or else). Our investigation is carried out incrementally, in terms of complexity of the models considered:

1. Infogan: Fully unsupervised with InfoGAN
2. Label-aware: InfoGAN with unsupervised G but D aware of labels
3. R-oriented: ‘Relational’ oriented learning

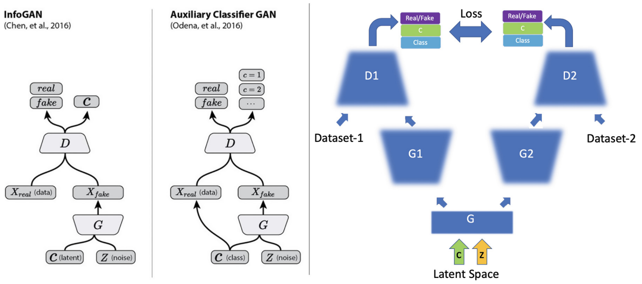

Fig. 2. Pictorial representations of InfoGAN, CatGAN, and R-oriented model (InfoGAN and CatGAN pictures from [17]).

We conduct many simulations with each model, with different latent code dimensionalities (that is c in the equations; the uncompressible noise z is always set to 32), for both the Categorical variable (number of possible choices) and the number of continuous Uniform[-1,1] variables in c. We also experiment changing the numerosity of the dataset from 1 to a maximum (2,3,4,...8), with the aim to study in each case the latent space with the methodology clarified at the beginning of this section, and compare qualitatively the degree of disentanglement reached in each case. For brevity, in this report, we provide illustrations only of specific but (we believe) representative cases of the results.

Infogan - Fully Unsupervised with InfoGAN: the model is the same as the one described in [4] with the exception of the latent space configuration, which for us is a field of exploration, and a change in the activation function in the first convolutional layer of the Discriminator, using the Absolute Value Rectification which has been shown [9] to be well suited for features that are invariant under a polarity reversal, like in our dataset. Few other modifications are also attempted (reducing the number of feature maps in the convolutional layers) aimed at simplifying the Generator and Discriminator networks, as the visual patterns in our synthetic data are simpler than the ones in the MNIST dataset. However, no significant differences were found in the overall results, both in terms of graphical reconstruction and, more importantly, with respect to numerosity disentanglement. In [4] it is shown how to derive a variational lower bound to the Mutual Information between the coding part of the latent “noise” and the output of the Generator $G(c, z)$. The minmax game is then re-defined as follows:

$$\min_{G,Q}\max_{D} V_{InfoGAN}(D,G,Q) \tag{6}$$

$$V_{InfoGAN}(D,G,Q) = V(D,G) - \lambda L_I(G,Q) \tag{7}$$

where D can intervene on L_I as Q in practice is often implemented using some layers of D. The L_I lower bound is shown in the same paper to be:

$$L_I(G,Q) = \mathbb{E}_{c' \backsim P(c'),(c,z)\backsim P(c,z)} log \hat{P}(c|G(c,z)) + H(c) \tag{8}$$

where $\hat{P}$ is the probability distribution estimated by Q. So, D and G aim at minimizing respectively:

$$\begin{aligned} D_{tot.loss} = -[\mathbb{E}_{x\backsim P_{data}}[\log(D(x))] + \mathbb{E}_{(c,z)\backsim P(c,z)}[\log(1 - D(G(c,z)))]] \\ - \lambda L_I(c; G(c,z)) \end{aligned} \tag{9}$$

$$G_{tot.loss} = \mathbb{E}_{(c,z)\backsim P(c,z)}[-\log(D(G(c,z)))] - \lambda L_I(c; G(c,z)) \tag{10}$$

Label-Aware - Unsupervised G but D Aware of Labels: This model is trained in a semi-supervised way and it could be considered as an union of the models presented in [4] and [11], as it extends the InfoGAN settings with an additional classifier whose output corresponds to the numerosity of the image at its input, and it introduces a regularization component to the total cost function in the following way: during the Discriminator training we expect the output of the classifier to be the correct class of the real data being submitted, while the output of the classifier is ignored when the generated data is passed to the Discriminator. During the Generator training, conversely, the output of the Generator -in every single instance in the batch- is expected to be classified with very low entropy (that is $P(y|G(z))$ expected to have very low entropy), and at batch level we expect the entropy of the classifier output to be high (that is, $P(y)$ expected to have very high entropy). Based on these considerations, it is possible to add a cost component (IS^{-1} in the following equations) to the total loss function for the Generator. This approach is similar to the reasoning behind the "Inception Score" metric for generative models, proposed in [12], and the whole setup used in this case has also similarities to the one proposed in [11]. In mathematical terms this translates in the following cost functions for D and G.

$$\begin{aligned} D_{tot.loss} = -[\mathbb{E}_{x\backsim P_{data}}[\log(D(x))] + \mathbb{E}_{(c,z)\backsim P(c,z)}[\log(1 - D(G(c,z)))]] \\ - \lambda L_I(c; G(c,z)) + CE(labels, y) \end{aligned} \tag{11}$$

$$\begin{aligned} G_{tot.loss} = \mathbb{E}_{(c,z)\backsim P(c,z)}[-\log(D(G(c,z)))] - \lambda L_I(c; G(c,z)) \\ + IS^{-1}(P(y|x), P(y)) \end{aligned} \tag{12}$$

where CE is the Cross Entropy and IS^{-1} is as just described[1].

[1] It is worth clarifying that for *each* component of the cost functions shown in all the equations, for all the three models considered, we apply a weighting hyper-parameter (thus, not only for the Information based reguliarized of the InfoGAN model), and we investigate empirically the effect of changing them.

R-oriented - Relational Oriented Learning: This model is trained in a semi-supervised way and it is based on the idea that in order to learn a disentangled representation of numerosity, we need to abstract it from the specific graphical appearance of an image. We thus force the model to represent features that are shared between two datasets, but expressed through different graphical representations, and whose informative content is what we want the model to learn to represent. This can be thought of as learning to represent a relation between sets rather than the features that one object (image) has or a set of objects (image dataset) exposes in statistical terms. This approach is inspired by the fact that a natural number can be defined as an equivalence class of finite sets under the equivalence relation of equinumerosity. We thus force the Generator to be multi-output over *the same latent space*. This in practice means that the whole setup described in Label-aware model is duplicated, sharing a bottom layer for the Generators which use *the same single latent space*. In this new duplicated setup we also add a component to the Generator loss function, the Jensen-Shannon divergence, which is symmetric with respect to its arguments, and we calculate it between the probability distributions over the output class predicted by the two classifiers. This has the goal to force to learn a probability distribution in the bottom layers of the Generator which must contain the necessary information to make both the reconstructions possible, while representing the same numerosity[2]. A pictorial representation of the model is shown in Fig. 2. In this setup the total loss functions used are slightly more complex than in the previous cases, as we have split the Generator in two lines of generation, rooted on the same latent space. With JS being the Jensen-Shannon divergence as per above, and with all the remaining quantities averaged between the two lines of the model, D and G aim at minimize respectively:

$$\begin{aligned} D_{tot.loss} = -[\mathbb{E}_{x \backsim P_{data}}[\log(D(x))] + \mathbb{E}_{(c,z)\backsim P(c,z)}[\log(1 - D(G(c,z)))]] \\ - \lambda L_I(c; G(c,z)) + CE(labels, y) \end{aligned} \tag{13}$$

$$\begin{aligned} G_{tot.loss} = \mathbb{E}_{(c,z)\backsim P(c,z)}[-\log(D(G(c,z)))] - \lambda L_I(c; G(c,z)) \\ + IS^{-1}(P(y|x), P(y)) + JS(P_1(y), P_2(y)) \end{aligned} \tag{14}$$

4 Experiments and Results

4.1 InfoGAN

With this model, we do not observe a clear "departure" from graphical features, in the sense that the study of the latent space always shows a strong connection with the graphical appearance of the images being generated, and very weak

[2] In our first setup to investigate this model we used, as second dataset, the labels themselves, feeding one line of the model with the labels and the other line with images. It must be noted however that this approach can be extended to a setup that does not use labels at all, however we leave this for future developments.

control over numerosity results from changes of the latent code dimensions. We note that whenever the dimensionality of the Categorical space is set to 2 (so with possible values (1,0) or (0,1)), the model appears to be capable of associating the Categorical part of the code to the kind of background/foreground of the image being generated (white over black versus black over white), but still not in a totally independent way, as changing from (1,0) to (0,1) it usually changes also the numerosity being represented (columns in the upper half, left in Fig. 3). Apart from that, no evidence of a correlation with numerosity, rather than with graphical features, is detected in any of the latent variables in all the tested options.

We provide latent space visualizations for two cases that we believe are representative of what we observe with this model. The first case (upper half of Fig. 3) has the latent code c configured as 2D Categorical variable and 1D continuous Uniform variable, with dataset numerosity equal to 2, abbreviated to (2D-1C-1,2); in this case the learning process cause the Categorical latent to represent the relation background/foreground of the image (columns on the upper left of Fig. 3), while the continuous variable shows a weak tendency to represent the numerosity contained in the dataset (1 or 2), only occasionally changing the numerosity displayed when it moves from negative values (left side of the rows in the upper part of Fig. 3) to positive values (right side of the rows).

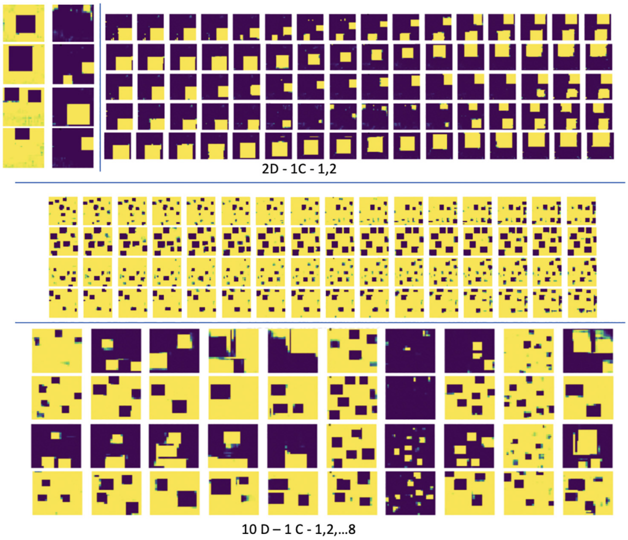

Fig. 3. Example of latent space exploration with the InfoGAN model, (2D-1C-1,2) upper part, (10D-1C-1,2,...8) lower part.

The second case (lower half Fig. 3), abbreviated as (10D-1C-1,2,...8) shows no evident signs of numerosity control in the continuous variable (rows in the center of Fig. 3) and the 10-dimensional Categorical variable does not seem to have picked any particular role (columns at the bottom of Fig. 3). We notice an increasing difficulty of the model to deal with the task as we increase the numerosity of the dataset, but we still obtain acceptable graphical quality of the generated samples. More interesting, we observe that a greater dimensionality of the Categorical part of the latent code never corresponds[3] to numerosity as learned descriptive (disentangled) dimension of the data.

4.2 Label-Aware

Similar results are obtained with the Label-aware model; we provide illustration for the same two cases considered in the previous section. We note a deterioration in the quality of the reconstructions contrasted by a slightly increase in the consistency with numerosity, as being represented in the Categorical part of the code (Fig. 5). As found for the previous model, in the case of binary Categorical latent variable, (2D-1C-1,2) in Fig. 5, we can see again that the Categorical variable picks the type of background while the continuous part of the latent space models some aspects that are fully connected to graphical features of the dataset. Generally, we notice that as we increase the numerosity of the input also this model fails to show any possible correlation between any latent code variable and numerosity.

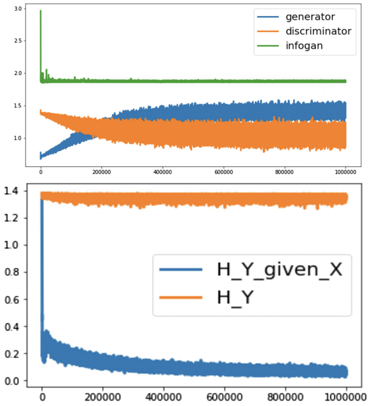

Fig. 4. Typical Losses and Entropies in Label-aware and R-oriented.

4.3 R-oriented

This model is designed to take as input two datasets, both with numerosity as either a strong or weak factor of variation. Here, we take a first exploratory step and we choose the inputs to be the dataset of images for one generative line, whereas the other generative line takes in, at the corresponding generator, the set of labels represented in one-hot encoding. This means that the two Generation lines are now expected to generate credible images on one side and valid coding for labels on the other, with *a priori* no relation since the Generators are never exposed to the association between labels and images from the dataset. For this model too, we provide latent space visualizations for the (2D-1C-1,2) case, as well as for (3D-1C-1,2,3), (4D-1C,1,2,3,4) and (6D-1C-1,2,...6), and we show that with this model, in all cases, numerosity is mapped to the Categorical variable

[3] Even when the Categorical dimensionality is somehow compatible with the numerosity being analized, for example with numerosity 5 and Categorical dimension 5, or Categorical dimension 10 to account for 8 quantities and 2 possible graphical expressions, w/b or b/w.

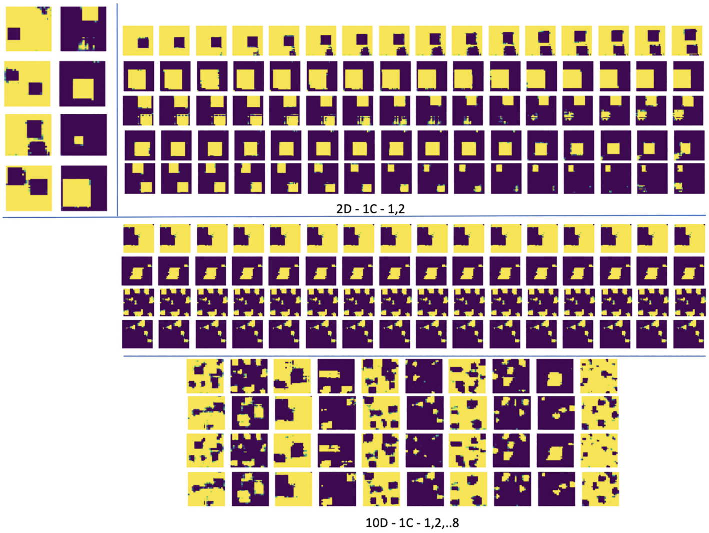

Fig. 5. Latent space exploration with Label-aware model with 2-dim. Categorical and 1-dim. continuous code; numerosity 1,2 (2D-1C-1,2); and 10-dim. Categorical and 1-dim. continuous, numerosity 1,2,...8 (10D-1C-1,2...8)

(Fig. 6) and that the continuous code only changes graphical aspects of the image being generated. Each picture shows the generated output image from the image generative line, while each digit on top each picture is the output of the classifier on the label generative line.

When setting the Categorical variable to a fixed value and changing randomly the remaining part the latent code, image numerosity (number of coherent colored areas in the image) stays the same in most of the cases, while the pattern of pixel activation of the image changes, giving rise to different ways of expressing the same numerosity. We note however that the graphical quality of the reconstruction is deteriorated compared to previous models; this might be connected to the setup used, in which the full latent code (c, z) is shared between the generative lines, even the uncompressible noise z, which seems neither necessary nor helpful. An appropriate calibration of the weights of the various cost functions components might also help improving the graphical reconstruction. However, our main goal was to investigate disentanglement, thus we accept a deterioration in graphical appearance leaving to future work taking care of this improvements.

5 Discussion

From our experiments it turned out, maybe not surprisingly, that it is not easy to map numerosity to any of the latent codes, either discrete or continuous, regardless of the dimensionality used in the various attempts. Being numerosity a concept of discrete nature, the first naive approach would be to model the

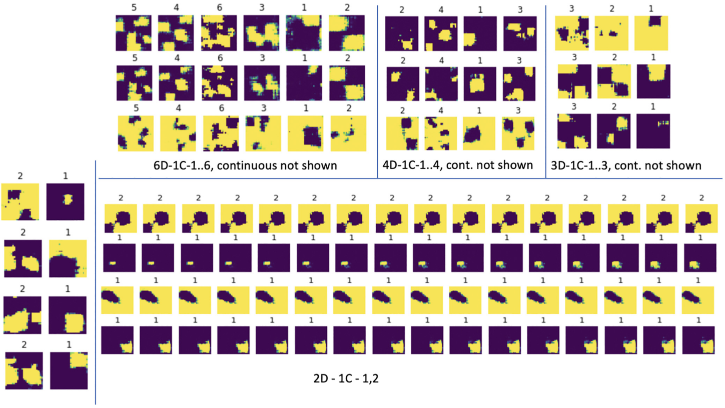

Fig. 6. Latent space study for the R-oriented model, for a (2D-1C-1,2), lower part of the image, for which we report a latent space study for both discrete and continuous variable, and, in the upper part, the latent space exploration with the Categorical variable for (6D-1C-1,...,6), (4D-1C-1,...,4), (3D-1C-1,...,3); the number above each single picture is the corresponding output of a classifier on the image generative line

InfoGAN code space as a Categorical code whose dimensionality is equal to the maximum number of objects present in the input images. This appeared to be effective with the first two models only with numerosity of 2 (see Sect. 4). From a semantic perspective, numerosity can be understood as an *inclusion* concept, for an image with 2 objects includes an image with 1 and so on; in this view one may think that the optimal representation for the concept might be discrete, but perhaps not simply Categorical. Similarly, representing numerosity using a continuous variable seems counter-intuitive, as the latent space should learn a distribution that is peaked around some values (that is, multimodal) to provide a “Categorical-like” representation of numerosity, along one continuous direction only. In this scenario, we might end up with parts of this continuous axis where the mapped images would be “morphing” from one numerosity to another one, leaving the result in this “transitory part of the latent space” undefined. The results we obtained confirm that learning a high level concept like numerosity in a generative model, in an unsupervised or semi-supervised fashion, and map it to an independent dimension of the latent space is not a straightforward task even for the InfoGAN model. This is in harmony with the findings in [8], and it may interpreted thinking about the semantically meaningful variations of “higher level of abstraction” as being overwhelmed by other statistical variations in the data, connected to much lower level features, like purely graphical for example. These variations must be somehow ignored by the learning process in order to let emerge the variations that carry the relevant semantic information. In this, the way the learning process is driven seems to be a way to leave the higher level semantic features to emerge, and we proposed here a possible approach, the R-oriented model, which is also inspired by the findings in [15].

6 Conclusion and Future Work

In this work we verified whether a powerful latent space generative model could learn a representation of numerosity as a disentangled factor of variation of input images derived from the dataset used in a seminal deep learning model of numerosity perception [1]. We tested several extensions of InfoGAN, mixing ideas coming from other GAN architectures like CatGAN [11], finding results that are in agreement with previous work but at the same time adding details and ideas to the field. We also introduced an alternative way to attack the problem of learning a disentangled representation of numerosity, introducing an ad-hoc *R-oriented* model, for which we have reported some preliminary but encouraging results. We leave to future efforts the in-depth study of the possibilities and developments of the latter approach.

References

1. Stoianov, I., Zorzi, M.: Emergence of a 'visual number sense' in hierarchical generative models. Nat. NeurosclI. **15**(2), 194–196 (2012)
2. Feigenson, L., Dehaene, S., Spelke, E.: Core systems of number. Trends Cogn. Sci. **8**(7), 307–314 (2004)
3. Zorzi, M., Testolin, A.: An emergentist perspective on the origin of number sense. Philos. Trans. Royal Soc. B Biol. Sci. **373**(1740) (2018)
4. Chen, X., Duan, Y., Houthooft, R., Schulman, J., Sutskever, I., Abbeel, P.: InfoGAN: Interpretable Representation Learning by Information Maximizing Generative Adversarial Nets (2016), arXiv:1606.03657
5. Wu, X., Zhang, X., Shu,X.: Cognitive Deficit of Deep Learning in Numerosity (2018), arXiv:1802.05160
6. Chen, S.Y., Zhou, Z., Fang, M., McClelland, J.L.: Can Generic Neural Networks Estimate Numerosity Like Humans? (2014)
7. Locatello, F., Bauer, S., Lucic, M., Gelly, S., Schölkopf, B., Bachem, O.: Challenging Common Assumptions in the Unsupervised Learning of Disentangled Representations (2018), arXiv:1811.12359
8. Zhao, S., Ren, H., Yuan, A., Song, J., Goodman, N., Ermon, S.: Bias and Generalization in Deep Generative Models: An Empirical Study arXiv:1811.03259v1 (2018)
9. Goodfellow, I., Bengio, Y., Courville, A.: Deep Learning. MIT Press, Cambridge (2016)
10. Goodfellow, I., et al.: Generative Adversarial Networks (2014), arXiv:1406.2661
11. Springenberg,J.: Unsupervised and Semi-supervised Learning with Categorical Generative Adversarial Networks (2015), arXiv:1511.06390
12. Salimans, T., Goodfellow, I., Zaremba, W., Cheung, V., Radford, A., Chen, X.: Improved Techniques for Training GANs (2016), arXiv:1606.03498
13. Barratt, S., Sharma, R.: A Note on the Inception Score (2018), arXiv:1801.01973
14. Katrina E., Drozdov, A.: Understanding Mutual Information and its Use in InfoGAN (2016)
15. Hill, F., Santoro, A., Barrett, D., Morcos, A., Lillicrap,T.: Learning to make analogies by contrasting abstract relational structure (2019), arXiv:1902.00120v1
16. Liu, Z., Luo, P., Wang, X., Tang, X.: Deep learning face attributes in the wild. In: ICCV (2015)
17. https://github.com/lukedeo/keras-acgan/blob/master/acgan-analysis.ipynb