In the supplementary material, we provide our analysis on AWA2, CUB and SUN datasets. We evaluate for the need of learning an optimal discriminative semantic aligned space in step 1 for both the inductive and transductive settings. Further, we examine the unseen class generator learned without any conditional information of seen classes in the transductive setting. In the end, we also provide details of the fine-tuned hyper-parameters utilized by SABR on the three benchamark datasets.

Type	**Method**	**AWA2**	**CUB**	**SUN**
I	SABR-I (DS)	65.1	59.9	61.2
	SABR-I (SR)	63.1	63.1	61.6
	SABR-I	**65.2**	**63.9**	**62.8**
T	SABR-T (DS)	88.0	70.1	65.1
	SABR-T (SR)	88.2	72.9	66.9
	SABR-T	**89.0**	**74.0**	**67.5**

[1] I - inductive ZSL setting, T - transductive ZSl setting.

Table 1: Performance in the conventional ZSL setting of different baselines with our proposed approach.

Type	**Method**	**AWA2**			**CUB**			**SUN**		
		MCA_u	MCA_s	H	MCA_u	MCA_s	H	MCA_u	MCA_s	H
I	SABR-I (DS)	24.9	90.3	39.0	43.3	60.2	50.3	45.4	32.5	37.9
	SABR-I (SR)	28.5	88.9	43.2	53.4	55.6	54.5	44.3	34.84	39.01
	SABR-I	30.3	93.9	**46.9**	55.0	58.7	**56.8**	50.7	35.1	**41.5**
T	SABR-T (DS)	80.7	88.3	84.4	64.2	68.1	66.1	55.8	38.6	45.6
	SABR-T (SR)	79.3	88.2	83.5	66.4	71.0	68.6	55.8	40.1	46.7
	SABR-T	79.7	91.0	**85.0**	67.2	73.7	**70.3**	58.8	41.5	**48.6**

[1] I - inductive GZSL setting, T - transductive GZSL setting

Table 2: Performance in the Generalized ZSL Setting of different baselines with our proposed approach.

1. Effect of learning the optimal latent space

In this section, we experimentally show that having both the classifier f_c and the regressor f_r to learn a discriminative and semantic aligned space helps to improve the performance in both the ZSL and GZSL setting. We provide two baselines to the existing SABR-I and SABR-T, to convey the importance of both the losses $\mathcal{L}_c$ and $\mathcal{L}_s$ defined for learning an optimal latent space. In the first baseline DS, we learn the latent space using just the classification loss, $\mathcal{L}_c$ i.e. making the latent space discriminative but not preserving the semantic relations. In the second baseline SR, we learn the latent space by preserving the semantic relations and using the similarity based cross entropy loss, $\mathcal{L}_s$. Once this latent space ψ is learned by either of the baselines, we model the generator to synthesize unseen class representations in both the inductive and transductive setting. Table 1 and 2 presents the results of the baselines with our approach on the three benchmark datasets in both ZSL and GZSL setting.

Firstly, the differences in the performance of the two baselines is more evident in the generalized ZSL setting than conventional ZSL setting. We hypothesize that the synthetic unseen class instances are confused with the seen classes in the final classifier leading to more notable differences in the performance. Secondly, we almost observe the trend that SABR-x (DS) $<$ SABR-x (SR) $<$ SABR-x where SABR-x can be SABR-I or SABR-T. This can be accounted due to the fact that the latent space learned by baseline DS encodes the discriminative features for seen classes but fails to generalize on unseen classes as no relationship is captured within the class embeddings. Although, if the semantic relations are captured by the latent space as in the baseline SR, it may suffer from hubness problem leading to reduced performance. Thus, SABR-x preserves the semantic relations between classes while discriminating the information among classes for an improved performance in both the conventional and generalized setting for the inductive and transductive approach.

Dataset	ZSL	GZSL		
		MCA_u	MCA_s	H
AWA2	12.6	11.4	91.2	20.2
CUB	9.3	8.8	21.1	12.4
SUN	1.7	1.3	15.2	2.4

Table 3: Performance in the conventional and generalized ZSL when unseen class instances are utilized without transfer of knowledge from seen classes

2. Learning marginals of unseen classes without conditionals

In this experiment, we learn a generative adversarial network which minimizes the marginal distribution between unseen class instances and generated instances for the transductive setting. This is done without utilizing any information of the seen class conditional data and thus there is no transfer of information from seen classes to unseen classes. For this, ω is equated to zero in equation 9.

The results are reported on AWA2, CUB and SUN in table 3. It can be clearly seen that learning the marginal distribution does not guarantee the alignment of conditionals of unseen classes, thus the model performs poorly in both ZSL and GZSL setting. This motivates us to learn a generator which minimizes marginal distribution by utilizing the conditional information of seen classes controlled by the hyper-parameter ω.

3. Hyper-parameter details

3.1. Learning the optimal latent space

In the first step of SABR-I and SABR-T, we learn a discriminative semantically aligned latent space, Ψ from the seen classes data. The dimension of this latent space is 1024 across all the datasets. Table 4 presents the batch size, weighing factor of semantic alignment (γ) in equation 3 and learning rate (lr).

Dataset	Latent_dim	Batch_size	γ	Lr
AWA2	1024	64	0.01	0.001
CUB	1024	64	0.01	0.001
SUN	1024	64	0.01	0.001

Table 4: Hyper-parameters of step-1, learning an optimal space

3.2. Bias Reducing Generator Network for SABR-I

In the inductive setting, we learn a generator $G^s : \langle z, c(y^s)\rangle \rightarrow \Psi$ where z denotes the noise vector , $c(y^s)$ is the class embedding of the seen classes and Ψ is the 1024 dimensional latent representations of the seen classes that we want the generator to synthesize. Table 5 gives the details of learning rate (lr), batch size, dimension of noise vector z (z_dim), dimension of class embedding (c(y)_dim), early stopping notch (stop), gradient penalty coefficient (λ) and β defined in equation 6.

3.3. Bias Reducing Generator Network for SABR-T

Given the marginal distribution of unseen classes, we learn a generator $G^u : \langle z, c(y^u)\rangle \rightarrow \Psi$ that takes noise z and semantic vector $c(y^u)$ as the input and outputs a synthetic representation of unseen class. This is done by transfer of the class conditional information from the seen classes to unseen classes. Hyper-parameters across the three datasets are depicted in table 6 i.e. learning rate (lr), batch size, dimension of noise vector z (z_dim), dimension of class embedding (c(y)_dim), gradient penalty coefficient (λ) and regularization parameter (ω).

Dataset	**Lr**	**Batch_size**	**z_dim**	**c(y)_dim**	**stop**	β	$\boldsymbol{\lambda}$
AWA2	0.00001	128	85	85	40	0.1	10
CUB	0.0002	128	312	312	45	0.1	10
SUN	0.0001	128	102	102	95	0.1	10

Table 5: Hyper-parameters for SABR-I

Dataset	**Lr**	**Batch_size**	**z_dim**	**c(y)_dim**	$\boldsymbol{\lambda}$	ω
AWA2	0.0001	128	85	85	10	0.008
CUB	0.0002	128	312	312	10	0.002
SUN	0.0001	128	102	102	10	0.002

Table 6: Hyper-parameters for SABR-T