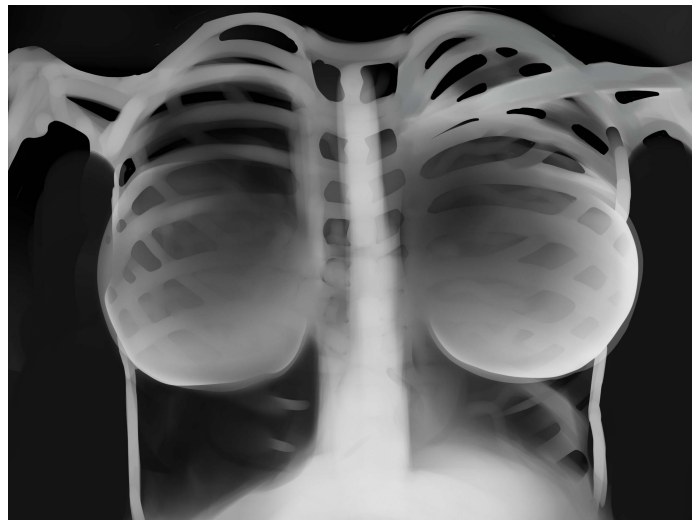




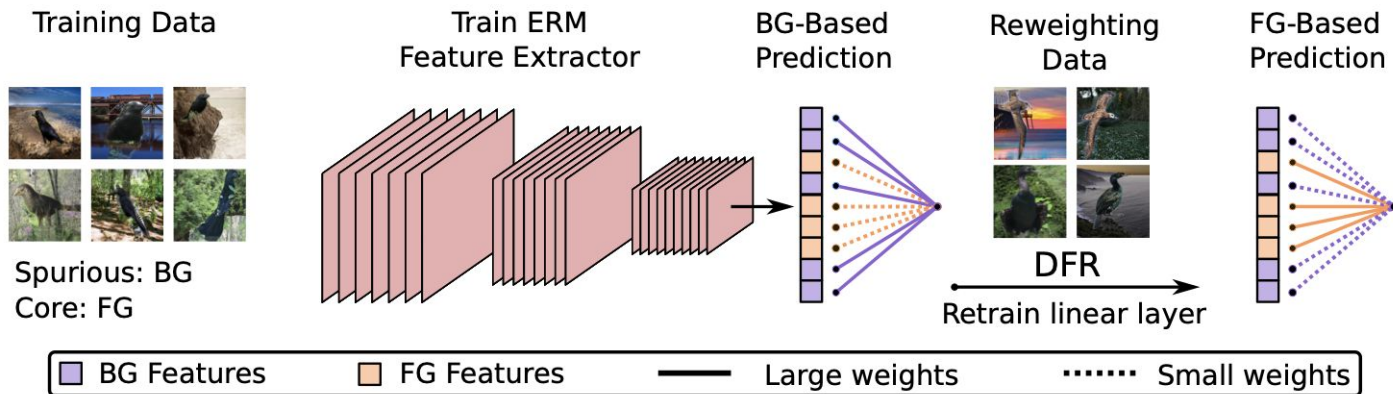
# **LAST LAYER RE-TRAINING IS SUFFICIENT FOR ROBUSTNESS TO SPURIOUS CORRELATIONS**

Elnikov Vlad

  
**Для чего это статья? - для борьбы с плохими  
корреляциями**



# Предложенный метод решения





## Цели данной статьи

1. Исследование влияния шумовых корреляций на работу нейронных сетей и оценка их влияния на точность классификации.
2. Проверка гипотезы о том, что переобучение только последнего слоя нейронной сети может быть достаточно для улучшения устойчивости модели к шумовым корреляциям.
3. Экспериментальное подтверждение эффективности метода переобучения последнего слоя на различных наборах данных и задачах классификации.
4. Сравнение результатов метода переобучения последнего слоя с другими методами улучшения устойчивости модели к шумовым корреляциям.



## Метод обучения (DFR)

1. **Инициализация модели:** сначала модель обучается на обучающем наборе данных.
2. **Оценка качества модели:** оценивается качество модели на тестовом наборе данных.
3. **Переобучение последнего слоя:** далее происходит переобучение только последнего слоя модели на меньшем взвешенном наборе данных. Это позволяет модели адаптироваться к новым данным и фокусироваться на ключевых признаках, игнорируя шумовые корреляции.
4. **Оценка улучшения качества:** после переобучения модель снова оценивается на тестовом наборе данных.

06\_musical\_instrument



05\_insect



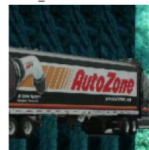
04\_carnivore



07\_primate



02\_wheeled vehicle



00\_dog



08\_fish



04\_carnivore



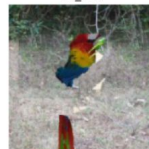
05\_insect



02\_wheeled vehicle



01\_bird



08\_fish



(a) IN-9 Original (Xiao et al., 2020) (b) IN-9 FG-Only (Xiao et al., 2020) (c) IN-9 Mixed-Rand (Xiao et al., 2020)

04\_carnivore



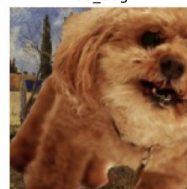
02\_wheeled vehicle



03\_reptile



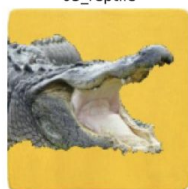
00\_dog



03\_reptile



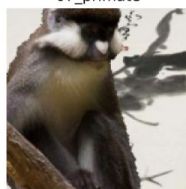
03\_reptile



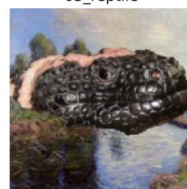
01\_bird



07\_primate



03\_reptile



01\_bird

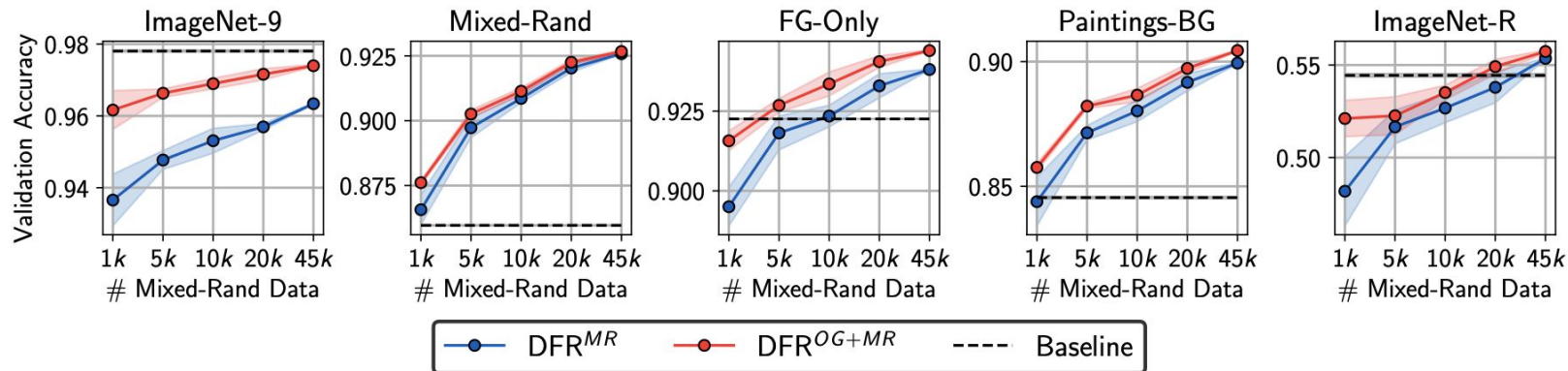


(f) ImageNet-9, Paintings-BG (ours)

# Результаты работы

Method	Group Info Train / Val	Waterbirds		CelebA		MultiNLI		CivilComments	
		Worst(%)	Mean(%)	Worst(%)	Mean(%)	Worst(%)	Mean(%)	Worst(%)	Mean(%)
JTT	✗ / ✓	86.7	93.3	81.1	88.0	72.6	78.6	69.3	91.1
CnC	✗ / ✓	88.5 $\pm$ 0.3	90.9 $\pm$ 0.1	88.8 $\pm$ 0.9	89.9 $\pm$ 0.5	-	-	68.9 $\pm$ 2.1	81.7 $\pm$ 0.5
SUBG	✓ / ✓	89.1 $\pm$ 1.1	-	85.6 $\pm$ 2.3	-	68.9 $\pm$ 0.8	-	-	-
SSA	✗ / ✓✓	89.0 $\pm$ 0.6	92.2 $\pm$ 0.9	<b>89.8</b> $\pm$ 1.3	92.8 $\pm$ 0.1	76.6 $\pm$ 0.7	79.9 $\pm$ 0.87	<b>69.9</b> $\pm$ 2	88.2 $\pm$ 2.
Group DRO	✓ / ✓	91.4	93.5	88.9	92.9	<b>77.7</b>	81.4	<b>69.9</b>	88.9
Base (ERM)	✗ / ✗	74.9 $\pm$ 2.4	98.1 $\pm$ 0.1	46.9 $\pm$ 2.8	95.3 $\pm$ 0	65.9 $\pm$ 0.3	82.8 $\pm$ 0.1	55.6 $\pm$ 0.6	92.1 $\pm$ 0.1
DFR <sub>Tr</sub> <sup>Val</sup>	✗ / ✓✓	<b>92.9</b> $\pm$ 0.2	94.2 $\pm$ 0.4	88.3 $\pm$ 1.1	91.3 $\pm$ 0.3	74.7 $\pm$ 0.7	82.1 $\pm$ 0.2	<b>70.1</b> $\pm$ 0.8	87.2 $\pm$ 0.3

# Результаты работы







## Выводы

- + Простота реализации: переобучение только последнего слоя требует меньше вычислительных ресурсов и времени
- + Эффективность: исследования показывают, что переобучение последнего слоя может значительно улучшить устойчивость модели
- + Меньшая вероятность переобучения
  - Ограниченность: переобучение только последнего слоя может быть недостаточным
  - Потеря информации: ограничение переобучения только последнего слоя может привести к потере информации
  - Не всегда эффективен