Применение синтетических данных, полученных с помощью генеративной нейросети, для повышения качества моделей детекции

Выпускная квалификационная работа бакалавра

Степанов Илья Дмитриевич
Научный руководитель: к.ф.-м.н. А. В. Грабовой
Научный констультант: А. В. Филатов

Кафедра интеллектуальных систем ФПМИ МФТИ
Специализация: Интеллектуальный анализ данных
Направление: 01.03.02 Прикладная математика и информатика

2025

Применение синтетических данных для детекции

Задача

Создание высококачественных аугментаций с помощью генеративной нейросети для повышения качества моделей детекции.

Проблема

Существующие методы генеративной аугментации для задачи детекции имеют недостатки: генерация объектов исходного класса; аугментация фона вместо самих объектов; аугментация изображений, адаптированная под конкретную прикладную задачу.

Цель

Разработать автоматизированный алгоритм, способный генерировать качественные аугментации и нивелировать недостатки существующих подходов. Провести сравнительный анализ влияния аугментаций и исследовать вклад отдельных компонентов метода.

Модель детекции

Рассмотрим модель детекции как отображение:

$$D_\omega : X \to \mathcal{F}(\hat{T}),$$

где X — множество изображений, $\hat{T}$ — пространство аннотаций для объектов, предсказанных моделью. $\mathcal{F}(\hat{T})$ — пространство предсказанных аннотаций изображений.

Пусть $\mathcal{L}(\omega)$ — функция потерь модели детекции. Решается следующая оптимизационная задача:

$$\omega^* = \arg\min_{\omega} \mathcal{L}(\omega),$$

Функция качества mAP

Рассмотрим функцию mAP (mean Average Precision):

$$\text{mAP} : \{\hat{T}\} \times \{T\} \times [0,1] \to [0,1],$$

Для каждого класса $c \in \mathcal{C}$ вычисляется функция AP (Average Precision):

$$\text{AP}(c, \tau, t, \hat{t}) = \int_0^1 P_c(r, \tau, t, \hat{t})\, dr,$$

где $P_c(r, \tau, t, \hat{t})$ — функция, задающая кривую Precision–Recall для класса c при пороге τ, $t \subseteq T$ — множество истинных аннотаций для класса c, $\hat{t} \subseteq \hat{T}$ — множество предсказанных аннотаций для класса c.

$$\text{mAP} = \frac{1}{|\mathcal{C}|} \sum_{c \in \mathcal{C}} \text{AP}(c, \tau, t, \hat{t}).$$

Функция качества $\text{mAP}_{50:95}$

Рассмотрим функцию $\text{mAP}_{50:95}$:

$$\text{mAP}_{50:95} : \{\hat{T}\} \times \{T\} \to [0, 1],$$

Определим промежуточную функцию $\text{AP}_{50:95}$ для каждого класса c:

$$\text{AP}_{50:95}(c,\ t,\ \hat{t}) \ = \ \frac{1}{10} \sum_{\tau \in \{0.50,\, 0.55,\, \ldots,\, 0.95\}} \text{AP}(c,\ \tau,\ t,\ \hat{t}).$$

$$\text{mAP}_{50:95} \ = \ \frac{1}{|\mathcal{C}|} \sum_{c \in \mathcal{C}} \text{AP}_{50:95}(c,\ t,\ \hat{t}).$$

Генеративная аугментация

Рассмотрим модель генеративной аугментации как отображение:

$$F_{\psi,\alpha,\beta,\gamma} : X \times [0,1] \longrightarrow (X_{\text{aug}} \times T_{\text{aug}}) \cup \{\varnothing\},$$

$$f_\psi : X \to M \times L \times T_{\text{aug}}$$

$$g_\alpha : X \times L \to P$$

$$h_\beta : X \times M \times P \to X_{\text{aug}}$$

$$r_\gamma : Y \times M \times L \times [0,1] \to \{0,1\}$$

где X — пространство изображений, X_{aug} — пространство аугментированных изображений, T_{aug} — пространство аннотаций аугментированных объектов, f_ψ — модель детекции объекта, g_α — модель генерации текстового запроса, h_β — модель генерации нового объекта, r_γ — модель фильтрации генераций, M — пространство масок объектов, P — пространство текстовых запросов, $L \subset P$ — пространство классов объектов.

Генеративная аугментация

$$F_{\psi,\alpha,\beta,\gamma}(x,\tau) = \begin{cases} (x_{\text{aug}},\ a_{\text{aug}}), & \text{если } r_\gamma\big(x_{\text{aug}},\ m,\ \ell,\ \tau\big) = 1, \\ \varnothing, & \text{если } r_\gamma\big(x_{\text{aug}},\ m,\ \ell,\ \tau\big) = 0, \end{cases}$$

$$\text{где } (m,\ \ell,\ a_{\text{aug}}) = f_\psi(x), \quad x_{\text{aug}} = h_\beta\big(x,\ m,\ g_\alpha(x,\ell)\big).$$

1. f_ψ извлекает маску и аннотацию объекта.
2. g_α формирует текстовый запрос для нового объекта на основе изначального класса и исходного изображения.
3. h_β генерирует аугментацию с помощью маски, текстового запроса и исходного изображения.
4. r_γ фильтрует некачественные аугментации с заданным порогом $\tau \in [0, 1]$.

Генеративная аугментация

Пусть $\mathcal{D} = \mathcal{D}_{\text{val}} \sqcup \mathcal{D}_{\text{train}}$. Рассмотрим аугментированный датасет для задачи детекции:

$$\mathcal{D}_{\text{aug}}(\tau) = \left\{ (x_i^{\text{aug}},\ t_i^{\text{aug}}),\ i = 1, \ldots, m \right\},$$

где $(x_i, t_i) \in \mathcal{D}_{\text{train}}$, $(x_i^{\text{aug}}, a_i^{\text{aug}}) = F_{\psi,\alpha,\beta,\gamma}(x_i, \tau)$, $a_i^* \in t_i$ — аннотация объекта с наибольшей площадью ограничивающего прямоугольника, $t_i^{\text{aug}} = \left(t_i \setminus \{ a_i^* \} \right) \cup \{ a_i^{\text{aug}} \}$ — аннотация аугментированного изображения, $\tau \in [0, 1]$ — пороговое значение для модели фильтрации.

Генеративная аугментация

Утверждение 1:
Пусть $\mathcal{D}_{\text{val}} = \{(x_i,\ t_i),\ i = 1, \ldots, k\}$. Существует такое значение $\tau^* \in [0, 1]$, что модели детекции f_{θ_1} и g_{ϕ_1}, обученные на объединённом датасете $\mathcal{D}_{\text{aug}}(\tau^*) \sqcup \mathcal{D}_{\text{train}}$, достигают не меньшего значения по функциям mAP_{50} и $\text{mAP}_{50:95}$ на $\mathcal{D}_{\text{val}}$, чем модели f_{θ_2} и g_{ϕ_2}, обученные на $\mathcal{D}_{\text{train}}$. То есть:

$$
\begin{aligned}
\text{mAP}_{50}\left(\{f_{\theta_1}(x_i)\}_{i=1}^k, \{t_i\}_{i=1}^k\right) &\geq \text{mAP}_{50}\left(\{f_{\theta_2}(x_i)\}_{i=1}^k, \{t_i\}_{i=1}^k\right), \\
\text{mAP}_{50:95}\left(\{f_{\theta_1}(x_i)\}_{i=1}^k, \{t_i\}_{i=1}^k\right) &\geq \text{mAP}_{50:95}\left(\{f_{\theta_2}(x_i)\}_{i=1}^k, \{t_i\}_{i=1}^k\right), \\
\text{mAP}_{50}\left(\{g_{\phi_1}(x_i)\}_{i=1}^k, \{t_i\}_{i=1}^k\right) &\geq \text{mAP}_{50}\left(\{g_{\phi_2}(x_i)\}_{i=1}^k, \{t_i\}_{i=1}^k\right), \\
\text{mAP}_{50:95}\left(\{g_{\phi_1}(x_i)\}_{i=1}^k, \{t_i\}_{i=1}^k\right) &\geq \text{mAP}_{50:95}\left(\{g_{\phi_2}(x_i)\}_{i=1}^k, \{t_i\}_{i=1}^k\right).
\end{aligned}
$$

Исследование влияния компонент: текстовый запрос

Рассмотрим модель аугментации следующего вида:

$$F'_{\psi,\beta,\gamma}(x,\tau) = \begin{cases} (x_{\text{aug}}, a_{\text{aug}}), & \text{если } r_\gamma\big(x_{\text{aug}},\, m,\, \ell,\, \tau\big) = 1, \\ \\ \varnothing, & \text{если } r_\gamma\big(x_{\text{aug}},\, m,\, \ell,\, \tau\big) = 0. \end{cases}$$

где $(m,\, \ell,\, a_{\text{aug}}) = f_\psi(x),\ x_{\text{aug}} = h_\beta\big(x,\, m,\, \ell\big)$.

Рассмотрим аугментированный датасет для задачи детекции:

$$\mathcal{D}'_{\text{aug}}(\tau) = \left\{ (x_i^{\text{aug}},\, t_i^{\text{aug}}),\ i = 1, \ldots, n \right\},$$

где $(x_i, t_i) \in \mathcal{D}_{\text{train}}$, $(x_i^{\text{aug}}, a_i^{\text{aug}}) = F'_{\psi,\beta,\gamma}(x_i, \tau)$, $a_i^* \in t_i$ — аннотация объекта с наибольшей площадью ограничивающего прямоугольника, $t_i^{\text{aug}} = \big(t_i \setminus \{\, a_i^* \,\}\big) \cup \{\, a_i^{\text{aug}} \,\}$ — аннотация аугментированного изображения, $\tau \in [0, 1]$ — пороговое значение для модели фильтрации.

Исследование влияния компонент: текстовый запрос

Утверждение 2:
Пусть $\mathcal{D}_{\text{val}} = \{(x_i,\ t_i),\ i = 1, \ldots, k\}$. Существует такое значение $\tau^* \in [0, 1]$, что модели детекции f_{θ_1} и g_{ϕ_1}, обученные на объединённом датасете $\mathcal{D}_{\text{aug}}(\tau^*) \sqcup \mathcal{D}_{\text{train}}$, достигают не меньшего значения по функциям mAP_{50} и $\text{mAP}_{50:95}$ на $\mathcal{D}_{\text{val}}$, чем модели f_{θ_2} и g_{ϕ_2}, обученные на $\mathcal{D}'_{\text{aug}}(\tau^*) \sqcup \mathcal{D}_{\text{train}}$. То есть:

$$
\begin{aligned}
\text{mAP}_{50}\left(\{f_{\theta_1}(x_i)\}_{i=1}^k, \{t_i\}_{i=1}^k\right) &\geq \text{mAP}_{50}\left(\{f_{\theta_2}(x_i)\}_{i=1}^k, \{t_i\}_{i=1}^k\right), \\
\text{mAP}_{50:95}\left(\{f_{\theta_1}(x_i)\}_{i=1}^k, \{t_i\}_{i=1}^k\right) &\geq \text{mAP}_{50:95}\left(\{f_{\theta_2}(x_i)\}_{i=1}^k, \{t_i\}_{i=1}^k\right), \\
\text{mAP}_{50}\left(\{g_{\phi_1}(x_i)\}_{i=1}^k, \{t_i\}_{i=1}^k\right) &\geq \text{mAP}_{50}\left(\{g_{\phi_2}(x_i)\}_{i=1}^k, \{t_i\}_{i=1}^k\right), \\
\text{mAP}_{50:95}\left(\{g_{\phi_1}(x_i)\}_{i=1}^k, \{t_i\}_{i=1}^k\right) &\geq \text{mAP}_{50:95}\left(\{g_{\phi_2}(x_i)\}_{i=1}^k, \{t_i\}_{i=1}^k\right).
\end{aligned}
$$

Исследование влияния компонент: фильтрация

Аналогично рассмотрим модель аугментации следующего вида:

$$F''_{\psi,\alpha,\beta}(x,\tau) \;=\; (x_{\text{aug}}, a_{\text{aug}})$$

где $(m,\, \ell,\, a_{\text{aug}}) = f_\psi(x),\; x_{\text{aug}} = h_\beta\big(x,\, m,\, g_\alpha(x,\ell)\big)$.

Рассмотрим аугментированный датасет для задачи детекции:

$$\mathcal{D}''_{\text{aug}}(\tau) = \left\{(x_i^{\text{aug}},\, t_i^{\text{aug}}),\; i = 1,\ldots,n\right\},$$

где $(x_i, t_i) \in \mathcal{D}_{\text{train}}$, $(x_i^{\text{aug}}, a_i^{\text{aug}}) = F''_{\psi,\alpha,\beta}(x_i,\tau)$, $a_i^* \in t_i$ — аннотация объекта с наибольшей площадью ограничивающего прямоугольника, $t_i^{\text{aug}} = \big(t_i \setminus \{\, a_i^* \,\}\big) \cup \{\, a_i^{\text{aug}} \,\}$ — аннотация аугментированного изображения, $\tau \in [0,1]$ — пороговое значение для модели фильтрации.

Исследование влияния компонент: фильтрация

Утверждение 3:
Пусть $\mathcal{D}_{\text{val}} = \{(x_i,\ t_i),\ i = 1, \ldots, k\}$. Существует такое значение $\tau^* \in [0, 1]$, что модели детекции f_{θ_1} и g_{ϕ_1}, обученные на объединённом датасете $\mathcal{D}_{\text{aug}}(\tau^*) \sqcup \mathcal{D}_{\text{train}}$, достигают не меньшего значения по функциям mAP_{50} и $\text{mAP}_{50:95}$ на $\mathcal{D}_{\text{val}}$, чем модели f_{θ_2} и g_{ϕ_2}, обученные на $\mathcal{D}''_{\text{aug}}(\tau^*) \sqcup \mathcal{D}_{\text{train}}$. То есть:

$$
\begin{aligned}
\text{mAP}_{50}\left(\{f_{\theta_1}(x_i)\}_{i=1}^k, \{t_i\}_{i=1}^k\right) &\geq \text{mAP}_{50}\left(\{f_{\theta_2}(x_i)\}_{i=1}^k, \{t_i\}_{i=1}^k\right), \\
\text{mAP}_{50:95}\left(\{f_{\theta_1}(x_i)\}_{i=1}^k, \{t_i\}_{i=1}^k\right) &\geq \text{mAP}_{50:95}\left(\{f_{\theta_2}(x_i)\}_{i=1}^k, \{t_i\}_{i=1}^k\right), \\
\text{mAP}_{50}\left(\{g_{\phi_1}(x_i)\}_{i=1}^k, \{t_i\}_{i=1}^k\right) &\geq \text{mAP}_{50}\left(\{g_{\phi_2}(x_i)\}_{i=1}^k, \{t_i\}_{i=1}^k\right), \\
\text{mAP}_{50:95}\left(\{g_{\phi_1}(x_i)\}_{i=1}^k, \{t_i\}_{i=1}^k\right) &\geq \text{mAP}_{50:95}\left(\{g_{\phi_2}(x_i)\}_{i=1}^k, \{t_i\}_{i=1}^k\right).
\end{aligned}
$$

Влияние аугментаций

Dataset	Model	Setting	Size	$\mathbf{mAP_{50}}$	$\mathbf{mAP_{50:95}}$
Pascal VOC	DETR	original	4000	57.2	41.2
		w/o expanded prompt	4000 + 4000	55.4	38.7
		w/o filter model	4000 + 4000	57.4	40.9
		ours	4000 + 4000	**58.2**	**41.4**
	YOLO	original	4000	59.6	41.5
		w/o expanded prompt	4000 + 4000	59.4	41.2
		w/o filter model	4000 + 4000	61.4	**43.2**
		ours	4000 + 4000	**61.5**	**43.2**
COCO	DETR	original	5000	26.6	17.6
		w/o expanded prompt	5000 + 5000	27.5	**17.8**
		w/o filter model	5000 + 5000	26	16.5
		ours	5000 + 5000	**27.8**	**17.8**
	YOLO	original	5000	26.7	17.4
		w/o expanded prompt	5000 + 5000	27.5	17.9
		w/o filter model	5000 + 5000	27.7	17.9
		ours	5000 + 5000	**28.2**	**18.3**

Проведение сравнительного анализа значений функций качества mAP_{50} и $\mathrm{mAP}_{50:95}$ моделей DETR и YOLO, обученных на датасетах Pascal VOC и COCO с применением аугментаций и без них, а также анализ влияния отдельных компонентов.

Выносится на защиту

1. Предложен автоматизированный подход к созданию аугментированных изображений.
2. Проведены эксперименты, демонстрирующие влияние аугментаций на качество работы модели детекции.
3. Проведён анализ влияния отдельных компонентов метода на итоговое значение функций качества.