ТРАНСПОРТНЫЕ СИСТЕМЫ
2.9.1 – Транспортные и транспортно-технологические системы страны, их регионов и городов, организация производства на транспорте;
2.9.4. – Управление процессами перевозок;
2.9.6 – Аэронавигация и эксплуатация авиационной техники;
2.9.8 – Интеллектуальные транспортные системы

УДК 004.946
DOI: 10.26467/2079-0619-2024-27-4-8-19

Нейросетевой подход к обеспечению визуальной когерентности в авиатренажерах дополненной реальности

А.Л. Горбунов[1], Ли Юньхань[1]
[1]Московский государственный технический университет гражданской авиации, г. Москва, Россия

Аннотация: В 2023 г. лидирующая авиакосмическая корпорация США Lockheed Martin объявила о разработке сразу нескольких основанных на технологиях расширенной/дополненной реальности (extended/augmented reality, XR/AR) тренажеров для пилотов TF-50, F-16, F-22 и F-35, отнюдь не являясь пионером в этом направлении – в 2022 г. аналогичные проекты запустили Boeing и ведущий британский производитель авиационной техники BAE Systems. В январе 2024 г. ВВС США инвестировали средства в разработку пилотских AR-симуляторов на основе смарт-очков дополненной реальности Microsoft Hololens, и тогда же компания Apple начала массовые продажи AR-гарнитуры Apple Vision Pro – трудно сомневаться в том, что в 2024 г. появится ряд новых авиатренажеров с применением этого устройства. Стремительное развитие нового поколения авиакосмической тренажерной техники – XR/AR-тренажеров – сопровождается бумом исследовательской активности в области визуальной когерентности (visual coherency, VC) сцен дополненной реальности: виртуальные объекты в этих сценах должны быть неотличимы от реальных. Именно VC обеспечивает новые возможности AR-тренажеров, принципиально отличающие их от ставших стандартными авиатренажеров с виртуальной реальностью. В последнее время VC все чаще обеспечивается нейросетевыми методами, при этом наиболее важными аспектами VC являются условия освещенности, поэтому основная доля исследований посвящена переносу этих условий (расположение источников света и их цветовой тон) из реального мира в виртуальный, но большинство известных подходов характеризуется неуниверсальностью и необходимостью выполнения ручных процедур. Данных недостатков не имеет основанный на двумерных спектральных преобразованиях изображений метод спектральной трансплантации, требующий, однако, определения размера трансплантируемой от реальной картины мира к виртуальному объекту части спектра. Настоящая статья посвящена разработке нейросетевой модели для механизма выбора оптимального размера спектрального трансплантата.

Ключевые слова: авиационные тренажеры, дополненная реальность, визуальная когерентность.

Для цитирования: Горбунов А.Л., Ли Ю. Нейросетевой подход к обеспечению визуальной когерентности в авиатренажерах дополненной реальности // Научный Вестник МГТУ ГА. 2024. Т. 27, № 4. С. 8–19. DOI: 10.26467/2079-0619-2024-27-4-8-19

Neural network approach to ensuring visual coherence in augmented reality flight simulators

A.L. Gorbunov[1], Yunhan Li[1]
[1]Moscow State Technical University of Civil Aviation, Moscow, Russia

Abstract: In 2023, the leading US aerospace corporation Lockheed Martin announced the simultaneous development of several extended/augmented reality (XR/AR) simulators for pilots of TF-50, F-16, F-22, and F-35 without being a pioneer in this area of

focus, in 2022 similar projects were launched by Boeing and the leading British aeronautical equipment manufacturer BAE Systems. In January 2024 the US Air Force invested in the development of pilot AR simulators based on Microsoft Hololens augmented reality smart glasses. At the same time, Apple began bulk sales of the Apple Vision Pro AR headset. It is difficult to doubt that in 2024 a variety of new aviation simulators will appear using this device. The rapid development of a new generation of aerospace simulator technology, i.e., XR/AR simulators, is accompanied by a boom in research in the field of visual coherence (VC) of augmented reality scenes: virtual objects in these scenes should be virtually identical with real ones. It is VC that provides new capabilities of AR simulators, which fundamentally distinguish from conventional flight simulators with virtual reality. Recently, VC has been increasingly provided by neural network methods, thereby, the most important aspects of VC are lighting conditions, so the major share of research is focused on transferring these conditions (location of light sources and their color tone) from the real world to the virtual one, but the great body of the known approaches are characterized by the lack of versatility and the need to perform manual procedures. These disadvantages are not found in the spectral transplantation method based on two-dimensional spectral image conversions, which, however, requires determining the size of the spectrum part being transplanted from the real picture of the world to a virtual object. This article is devoted to the development of a neural network model for the mechanism of selecting the optimal size of a spectral transplant.

Key words: flight simulators, augmented reality, visual coherence.

For citation: Gorbunov, A.L., Li, Yu. (2024). Neural network approach to ensuring visual coherence in augmented reality flight simulators. Civil Aviation High Technologies, vol. 27, no. 4, pp. 8–19. DOI: 10.26467/2079-0619-2024-27-4-8-19

Введение

В 2023 г. лидирующая авиакосмическая корпорация США Lockheed Martin объявила о разработке сразу нескольких основанных на технологиях расширенной реальности (extended reality, XR) тренажеров для пилотов TF-50, F-16, F-22 и F-35, отнюдь не являясь пионером в этом направлении, – в 2022 г. аналогичные XR-проекты запустили Boeing и ведущий британский производитель авиационной техники BAE Systems. Однако инициатива Lockheed Martin особо показательна широким спектром XR-симуляторов как в смысле множественности типов воздушных судов, так и разнообразия подвидов технологий XR: «обычная» дополненная реальность (augmented reality, AR) – проекты в кооперации с Korea Aerospace Industries и Red 6 Aerospace, а также комбинация виртуальной реальности (virtual reality, VR) и AR – проект на основе решения XTAL 3 от Vrgineers (интересный факт: подобный подход был описан в статье, опубликованной в «Научном Вестнике МГТУ ГА» еще в 2016 г. [1]). В январе 2024 г. ВВС США инвестировали средства в разработку пилотских AR-симуляторов компанией Microsoft на основе смарт-очков дополненной реальности Microsoft Hololens, и тогда же хорошо известная своевременностью своих реакций на запросы IT-рынка компания Apple начала массовые продажи AR-гарнитуры Apple Vision Pro – трудно сомневаться в том, что в ближайшем будущем появится ряд новых авиатренажеров с применением этого устройства.

Очень быстрое (даже по меркам IT-сферы) развитие нового поколения авиакосмической тренажерной техники – XR/AR-тренажеров – сопровождается всплеском исследовательской активности в области визуальной когерентности (visual coherence, VC) сцен дополненной реальности: виртуальные объекты в этих сценах должны быть неотличимы от реальных. Повышенный интерес исследователей обусловлен тем, что именно VC обеспечивает новые возможности AR-тренажеров, принципиально отличающие их от ставших де-факто стандартными современных авиатренажеров с VR. Все лидеры IT-отрасли (Microsoft, Google, Apple, Samsung, Sony и проч.) рассматривают AR как следующую (после появления смартфонов) «большую волну» революционных изменений в цифровой электронике, поэтому проблема VC становится ключевой и для IT в целом. Как следствие, зарубежная библиография по VC содержит уже более тысячи пунктов (тогда как в российском научном пространстве обнаруживаются только единичные публикации, e. g. [2]).

VC зависит от многих факторов: расположения источников света, воспроизведения виртуальных теней и цветового тона, взаим-

ных отражений реальных и виртуальных объектов, текстур поверхностей виртуальных объектов, оптических аберраций, конвергенции и аккомодации и др., но обычно наиболее важными являются условия освещенности, поэтому основная доля исследований посвящена переносу этих условий (расположение источников света и их цветовой тон) из реального мира в виртуальный. Практикуются два базовых варианта обеспечения VC, основанные: 1) на предварительном исследовании условий освещенности в реальном мире (проблемы: длительная и сложная процедура, требует специального оборудования) и 2) формировании представления об условиях освещенности на основе анализа изображений реального мира (проблемы: трудности извлечения информации об освещенности из реальных изображений приводят к неоднозначным результатам). В последнее время VC, как правило, обеспечивается методами искусственного интеллекта [3, 4].

Указанные проблемные моменты базовых методов VC приводят к их неуниверсальности, неавтоматическому характеру и невозможности передачи не только цветовой палитры, но и всех основных характеристик изображения с помощью одной процедуры. От данных недостатков свободен способ обеспечения VC, представленный в [2]. Способ ориентирован на характерные для авиационных приложений условия (сцены AR с реальными естественными ландшафтами и рассеянным освещением) и базируется на аппарате двумерных спектральных преобразований изображений. Предложена технология двумерной спектральной трансплантации, которая обеспечивает прямую передачу характеристик цвета, яркости и контраста от реального фона к виртуальным объектам: способ предполагает замену части спектра изображения виртуального объекта на такую же часть спектра изображения реального мира с последующим обратным преобразованием спектра с замененной частью в обычное изображение.

При этом возникает задача выбора оптимального размера трансплантируемой части спектра изображения реального мира (далее – трансплантат). По мере увеличения объема трансплантируемых пространственных частот они начинают содержать информацию о содержимом реального изображения, поэтому ограничение размера трансплантата необходимо для исключения гибридного эффекта. Сложность выбора этого ограничения обусловлена тем, что он связан с характером и параметрами изображения, часть спектра которого используется в качестве трансплантата. Механизм выбора должен обеспечивать: а) максимальную схожесть изображения виртуального объекта с картиной реального мира в смысле яркостной, контрастной и цветовой тонировки и б) отсутствие в изображении виртуального объекта «следов» содержания реальной картины. Характер реальной картины может быть самым разным, поэтому невозможно сконструировать строгий аналитический инструмент для решения данной задачи. Такой инструмент надо научить учитывать множество разных типов изображений, что естественным образом приводит к идее нейросетевой реализации. Нейросеть для определения оптимального размера трансплантата должна уметь детектировать признаки стиля и объекты в реальной компоненте сцены AR, что сегодня является одной из ключевых проблем компьютерного зрения.

Обзор публикаций

Обширный (230 источников) обзор нейросетевых методов детектирования объектов в изображениях содержится в статье [5], приведем несколько характерных примеров. Существуют два основных подхода к обнаружению объектов: снизу вверх (bottom-up, BU) [6] и сверху вниз (top-down, TD) [7]. Здесь «верх» и «низ» ассоциируются с высоким и низким уровнем семантики признаков, используемых при детектировании объектов. Контраст локальных объектов играет центральную роль в BU, независимо от семантического содержания сцены. Чтобы выявить контрастные границы объектов, из совокупности пикселей извлекаются различные локальные и глобальные объекты,

например грани [8]. Типовой нейросетью этой категории можно назвать R-CNN [9]. R-CNN использует выборочный поиск, чтобы сгенерировать около 2 тыс. предложений по фрагментам для изображения. Метод выборочного поиска основан на простой группировке по принципу BU и указателях значимости, что позволяет быстро получать уточненные списки фрагментов-кандидатов произвольных размеров и сокращать пространство поиска при обнаружении объектов.

Несмотря на очевидные достоинства сетей типа R-CNN, при их практическом применении для обнаружения объектов выявляется много недостатков:

- из-за наличия полносвязных слоев (все входные нейроны связаны со всеми выходными) требуется входное изображение фиксированного размера, что приводит к повторному пересчету всей сети для каждой оцениваемой области, занимая много времени в период тестирования;
- обучение R-CNN – многоступенчатый конвейер. Сначала создается сверточная сеть по фрагментам-кандидатам и выполняется точная настройка. Затем классификатор, работающий по известной модели векторного пространства, определяет вероятность наличия объектов. Наконец, обучаются регрессоры с ограничивающими фрагменты рамками;
- обучение требует больших затрат времени и ресурсов. Характеристики извлекаются из предложений по различным фрагментам и сохраняются на диске. Обработка относительно небольшого обучающего набора очень глубокими сетями, такими как VGG-16 (см. ниже), занимает много времени. Объем памяти, необходимый для этих процедур, также должен быть весьма значительным;
- хотя выборочный поиск может генерировать предложения по фрагментам при относительно большом количестве оценок, полученные предложения по фрагментам тем не менее являются избыточными.

Для решения этих проблем был предложен ряд методов. GOP [10] использует ускоренную сегментацию на основе геодезических данных. MCG [11] выполняет поиск в различных масштабах изображения множества иерархических сегментаций и комбинаторно группирует различные области для получения предложений. Вместо выделения визуально различимых сегментов метод граничных рамок [12] использует идею о том, что объекты скорее всего будут находиться в рамках с меньшим количеством контуров, определяющих их границы. Некоторые исследователи изменяют ранжирование или уточняют предварительно выделенные области, с тем чтобы исключить ненужные и получить ограниченное количество ценных – DeepBox [13] и SharpMask [14].

Однако высокоуровневая и многомасштабная семантическая информация не может быть задана с помощью этих низкоуровневых структур. Обнаружение выделяющихся объектов TD ориентировано на конкретные задачи и использует предварительные знания о категориях объектов (например, воздушные суда для авиаприложений) для управления созданием карт выделяемых объектов. К примеру, при семантической сегментации генерируется карта значимости для ассоциации наборов пикселей с определенными категориями объектов посредством подхода TD. Характерный для этого подхода образец – сеть YOLO, описанная в [15]. YOLO разбивает входное изображение на сетку ячеек, и для каждой ячейки вычисляется вероятность наличия объекта, содержащегося в ней. Ячейка сетки предсказывает ограничивающие ее рамки и соответствующие им оценки достоверности. TD можно рассматривать как механизм фокусировки внимания, который устраняет элементы, вероятность принадлежности которых объекту невысока.

В настоящее время оба подхода часто реализуются в виде вариаций на тему хорошо зарекомендовавшей себя и широко применяемой сверточной нейросети VGG-16 [16]. Еще один широкий обзор по теме детектирования объектов представлен в работе [17], где дается предварительный экскурс в историю развития и архитектуры соответствующих нейросетей.

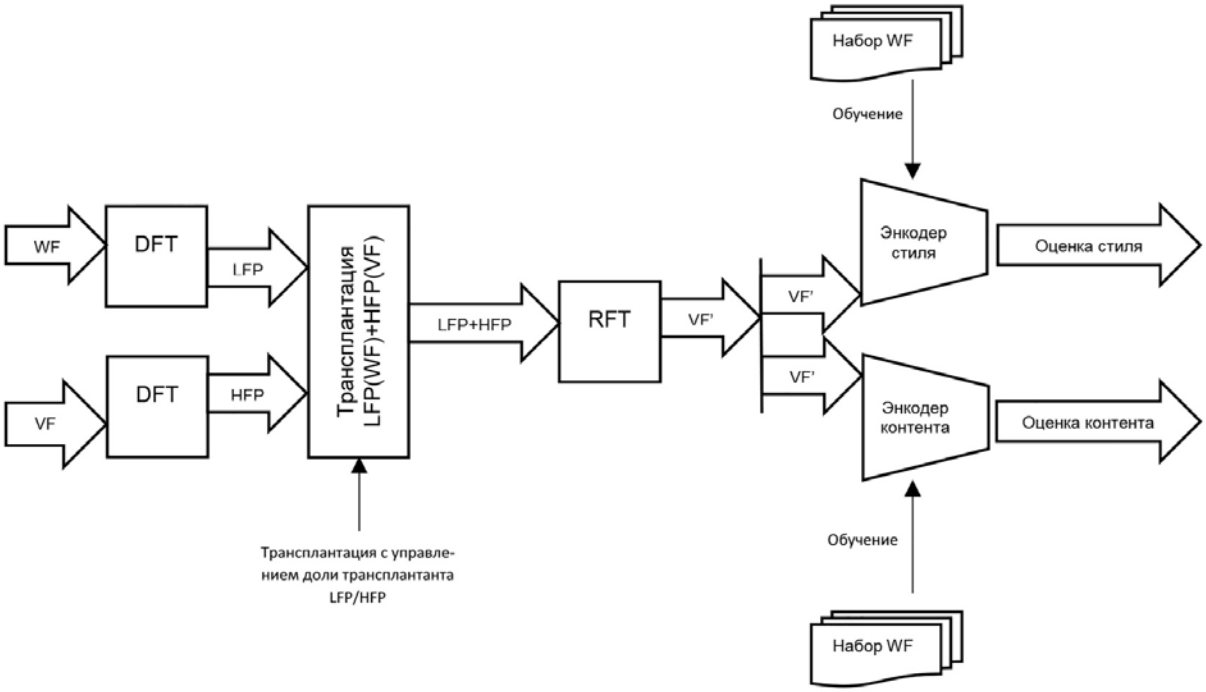

Рис. 1. Общая схема определения оптимального размера трансплантата
Fig. 1. General scheme for determining the optimal size of the transplant

Метод

Общая схема предлагаемого механизма определения оптимального размера трансплантата для метода спектральной трансплантации [2] показана на рис. 1.

На схеме: WF (world frame) – кадр картины реального мира; VF (virtual frame) – кадр картины виртуального мира; DFT (direct Fourier transform) – двумерное прямое преобразование Фурье; LFP (low frequency part) – низкочастотная часть спектра, полученного после DFT, трансплантат; HFP (high frequency part) – высокочастотная часть спектра, полученного после DFT, дополняющая LFP до полного спектра; RFT (reverse Fourier transform) – двумерное обратное преобразование Фурье; VF’ – кадр картины виртуального мира после трансплантации; энкодер стиля – сверточная нейросеть, обученная на массиве изображений WF и детектирующая признаки характерного для них стиля (особенности яркостной, контрастной и цветовой тонировки) в виде обобщенной количественной оценки выявления стиля; энкодер контента – сверточная нейросеть, обученная на массиве изображений WF и детектирующая признаки характерных для них объектов в виде обобщенной количественной оценки выявления контента.

На рис. 2 показаны результаты трансплантации с различными параметрами для различных типов виртуальных объектов – виртуальных моделей самолетов, отличающихся текстурой поверхности, маркировкой и блеском поверхности. Части рисунка *а*, *б* и *в* содержат виртуальный самолет со сложными текстурами, текстовыми обозначениями и отражениями виртуальных источников света; *г*, *д* и *е* содержат виртуальный самолет с простыми контрастными цветами. Части *а* и *г* содержат виртуальные объекты без трансплантации; *б* и *д* содержат виртуальные объекты после трансплантации LFP с тремя пространственными частотами; *в* и *е* содержат виртуальные объекты после трансплантации LFP с пятью пространственными частотами. Виртуальные объекты намеренно показаны без других эффектов VC (тени, локальное освещение и т. д.), чтобы продемонстрировать чистые результаты метода.

Рис. 2. Слева *а*, *г*: AR-сцены, состоящие из WF и VF без трансплантации LFP. Средний столбец *б*, *д*: AP-сцены, составленные после трансплантации LFP с тремя частотами. Справа *в*, *е*: AR-сцены, составленные после трансплантации LFP с пятью частотами

Fig. 2. On the left *а*, *г*: AR scenes consisting of WF and VF without the LFP transplantation. The middle column *б*, *д*: AR scenes composed after the LFP transplantation with three frequencies. On the right *в*, *е*: AR scenes composed after the LFP transplantation with five frequencies

Нейросеть VGGNet, производные формы которой целесообразно использовать в качестве энкодера в схеме на рис. 1, показала высокие результаты в известном конкурсе по компьютерному зрению ILSVRC, будучи обучена на наборе данных ImageNet (в настоящее время содержит более 5 миллионов размеченных высококачественных изображений, разделенных на более чем 22 тысячи категорий, среди которых воздушные суда) и доказав, что малоразмерные свертки плюс увеличение глубины сети могут ощутимо улучшить производительность. ImageNet представляет собой массив аннотированных картинок, это один из наиболее представительных в настоящее время открытых наборов данных в сфере компьютерного зрения. Каждое изображение в ImageNet содержит несколько выделенных вручную прямоугольных фрагментов, сопровождаемых проиндексированной информацией о содержании фрагментов, что позволяет обучить нейросеть распознаванию объектов требуемой категории. VGGNet является модернизированной версией сети AlexNet, в которой заменены большие фильтры (размера 11 и 5 в первом и втором сверточном слое соответственно) на несколько фильтров размера 3×3, следующих один за другим. Наиболее часто используемой версией VGGNet является VGG-16, реализация которой входит во многие открытые библиотеки нейросетевого программного обеспечения (например, TensorFlow от Google). Другая известная вариация на тему VGGNet – VGG-19. Различие между VGG-16 и VGG-19 состоит в том, что VGG-19 имеет на три сверточных слоя больше, чем VGG-16, на большинстве задач показывает лучшую

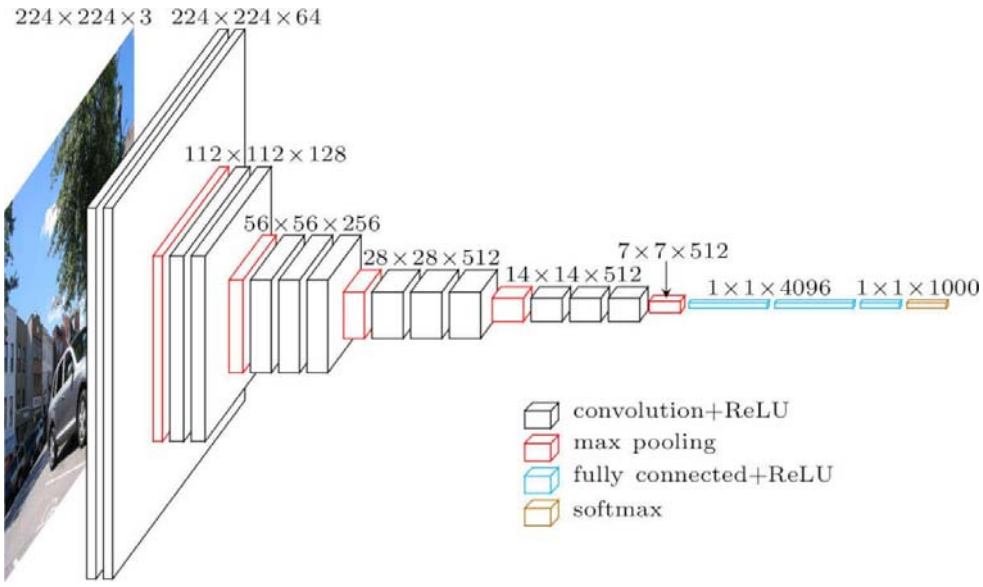

Рис. 3. Структура сети VGG-16 (иллюстрация neurohive.io)
Fig. 3.VGG-16 network structure (illustration from neurohive.io)

производительность, но требует больше вычислительных ресурсов. Структура VGG-16 показана на рис. 3.

В VGG-16 подается видеопоследовательность или изображение размерностью 224 × 224 × 3 (3 цветовых канала 224 × 224), серые (convolution) прямоугольные блоки представляют сверточные слои; красные (max pulling) – слои, выделяющие наиболее существенные признаки; синие (fully-connected) – полносвязные слои; желтый (softmax) – слой, используемый для вероятностного прогноза признаков, ReLU (Rectified linear unit) – применяемая функция активации нейронов, снижающая вычислительные затраты на обучение. Число 16 в VGG-16 обозначает количество обучаемых – сверточных и полносвязных – слоев сети VGG.

Использование VGG в качестве энкодеров для выявления признаков контента и стиля в изображении становится возможным за счет специальной конфигурации элементов фильтрующих матриц сверточных, синапсов полносвязных слоев и обработки результатов их вычислений. Для оценок контента и стиля сегодня применяются различные меры, обычно являющиеся развитием описанных в работе [18]. Например, для оценки контента предлагается использовать

$$L_C = \frac{1}{4n_C\, n_H n_W} \sum_{i=1}^{n_H} \sum_{j=1}^{n_W} \sum_{k=1}^{n_C} (p_{i,j,k}^{VF\prime} - t_{i,j,k}^{WF})^2 \quad , \tag{1}$$

здесь $\{p_{i,j,k}\}$ – набор признаков для VF’; $\{t_{i,j,k}\}$ – набор признаков для WF; n_H, n_W – высота и ширина карты признаков (для последнего слоя признаков на рис. 3 – 7 × 7); n_C – число каналов (на рис. 3 n_C = 512); множитель перед суммой используется для нормировки;

а для оценки стиля евклидово расстояние между матрицами Грама G на последних сверточных слоях для VF’и WF

$$J_S = \frac{1}{n_C^2 (n_H n_W)^2} \sum_{i=1}^{n_C} \sum_{j=1}^{n_c} (G_{i,j}^{VF\prime} - G_{i,j}^{WF})^2. \tag{2}$$

Приведенные оценки (1) и (2) для последних слоев имеют смысл различия между VF’ и WF. Задача определения оптимального размера трансплантата при этом формулируется как максимизация L_C при минимально возможном J_S либо минимизация J_S при максимально возможном L_C. В общем случае (применение предобученной сети VGG-16

Таблица 1
Table 1

Структура экспериментальной серии
Structure of the experimental series

Зависимые переменные	**Независимые переменные**			
	тип WF_1			**тип $WF_{2...L}$**
	тип VF_1		**тип $VF_{2...M}$**	
	LFP/HFP_1	**$LFP/HFP_{2...N}$**		
Оценка стиля				
Оценка контента				

необязательно) основанная на схеме рис. 1 процедура определения оптимального размера трансплантата подразумевает следующие шаги.

1. Настройка нейросетевых моделей энкодеров на выявление признаков стиля и содержания.

2. Обучение энкодеров на наборах WF, характерных для предполагаемых сцен дополненной реальности. При использовании VGG-16 сеть уже обучена на картинках с аннотированными фрагментами, содержащими изображения воздушных судов и строений.

3. Проведение серии экспериментов, каждый из которых представляет собой проход по схеме, приведенной на рис. 1. При этом в качестве независимых переменных могут использоваться соотношение LFP и HFP (управление долей трансплантата на рис. 1), тип WF и тип VF; в качестве независимых переменных могут быть использованы оценка контента (1) и оценка стиля (2) либо иные варианты расчета таких оценок. Структура экспериментальной серии иллюстрируется табл. 1.

Цель серии экспериментов – нахождение наиболее близкого расположения точек *max (оценка контента)* и *min (оценка стиля)* для выбранного диапазона LFP/HFP, которое определяет оптимальное соотношение LFP/HFP и соответствующий оптимальный размер трансплантата LFP. Необходимость экспериментирования с разными типами WF и VF (M > 1, L > 1 в табл. 1) диктуется спецификой предполагаемых сцен дополненной реальности. Шаг изменения соотношения LFP/HFP – один дискрет пространственной частоты, целесообразно использовать нижние 5 % всех возможных соотношений. Разумный предел N числа соотношений LFP/HFP, по ряду которых находится минимум для модуля разности нормализованных значений оценок контента и стиля, выявляется по затуханию флуктуаций данных оценок.

Осложняющим фактором при определении оптимального размера трансплантата LFP (как и при спектральной трансплантации, особенно при обработке больших изображений высокого разрешения) является большая вычислительная затратность механизма двумерного дискретного преобразования Фурье, которое необходимо выполнять трижды при каждом проходе по схеме на рис. 1. Снизить остроту проблемной ситуации может использование хорошо известного в аэронавигации алгоритма Волдера [19], обеспечивающего существенное уменьшение объема вычислений при спектральном преобразовании. Алгоритм Волдера представляет собой итерационную процедуру вращения вектора вокруг начала координат, при этом на каждой итерации осуществляется поворот на угол $arctg2^{-i}$, где i – номер итерации (реализуется так называемым CORDIC-процессором). То обстоятельство, что преобразование Фурье также имеет геометрическую интерпретацию в виде поворотов комплексных векторов, позволяет эффективно применять алгоритм Волдера при расчете, к примеру, быстрого преобразования Фурье. Конечное число итераций алгоритма Волдера (обычно несколько десятков) привносит дополнитель-

ную погрешность в результат, однако в случае обработки изображений психофизиологические особенности зрительного анализатора позволяют пренебречь таковыми.

Заключение

Происходящий в настоящий момент переход от современных авиатренажеров виртуальной реальности к XR/AR-тренажерам отражает общий текущий тренд IT-технологий, обусловленный рядом преимуществ AR по сравнению с VR. Эти преимущества связаны с параллельным сосуществованием в AR виртуальных и реальных объектов, среди них можно выделить следующие:

- реальное расширяет виртуальное – в VR сенсорный опыт пользователя преимущественно ограничен видео- и аудиоэффектами, тогда как в AR присутствует весь спектр ощущений реального мира;
- виртуальное расширяет реальное – в AR возможно моделирование ситуаций, которые невозможно или небезопасно создавать в реальном мире, оставаясь при этом в его рамках;
- естественный интерфейс – в AR управление виртуальными объектами движением зрачков и жестами предельно упрощает взаимодействие пользователя и компьютера;
- связь между виртуальными и реальными объектами – позволяет повысить дидактическую эффективность обучающих систем.

Неслучайно указанный переход начинается именно в авиационных приложениях, поскольку нынешние VR-авиатренажеры практически исчерпали свои возможности, а VC в сценах AR обещает реализацию комплекса принципиально новых направлений обучения.

Предложенная нейросетевая модель для определения оптимального размера трансплантата при спектральной трансплантации, осуществляемой с целью обеспечения VC, позволяет решить сложную задачу усреднения при выявлении признаков стиля и контента в чрезвычайно разнохарактерных изображениях, которые могут возникнуть в качестве реальной компоненты сцен дополненной реальности в авиакосмических обучающих системах новейшего поколения – тренажерах дополненной реальности. Реализация такого подхода расширит спектр XR-симуляторов (e. g., безопасное моделирование опасных ситуаций в реальной среде в учебных целях) и повысит их дидактическую эффективность.

Список литературы

1. **Горбунов А.Л.** Тренажер аэродромной спецтехники // Научный Вестник МГТУ ГА. 2016. № 225 (3). С. 92–97.
2. **Горбунов А.Л.** Визуальная когерентность в дополненной реальности // Advanced Engineering Research. 2023. № 2. С. 180–190. DOI: 10.23947/2687-1653-2023-23-2-180-190
3. **Tewari A., Fried O., Thies J. и др.** State of the art on neural rendering // Computer Graphics Forum. 2020. Vol. 39, iss. 2. Pp. 701–727. DOI: 10.1111/cgf.14022
4. **Einabadi F., Guillemaut J., Hilton A.** Deep neural models for illumination estimation and relighting: A survey // Computer Graphics Forum. 2021. Vol. 40, iss. 6. Pp. 315–331. DOI: 10.1111/cgf.14283
5. **Ghasemi Y.** Deep learning-based object detection in augmented reality: A systematic review / Y. Ghasemi, H. Jeong, S. Choi, J. Lee, K. Park [Электронный ресурс] // Computers in Industry. 2022. Vol. 139. ID: 103661. DOI: 10.1016/j.compind.2022.103661 (дата обращения: 05.01.2024).
6. **Tu W.-C., He Q., Chien S.-Y.** Real-time salient object detection with a minimum spanning tree // Proceedings of IEEE CVF Computer Vision and Pattern Recognition Conference (CVPR), 2016. Pp. 2334–2342.
7. **Yang J., Yang M.-H.** Top-down visual saliency via joint CRF and dictionary learning [Электронный ресурс] // Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. USA, RI, 2012. Pp. 2296–2303. DOI: 10.1109/CVPR.2012.6247940 (дата обращения: 05.01.2024).

8. Rosin P.L. A simple method for detecting salient regions // Pattern Recognition. 2009. Vol. 42, iss. 11. Pp. 2363–2371. DOI: 10.1016/j.patcog.2009.04.021

9. Girshick R. Rich feature hierarchies for accurate object detection and semantic segmentation / R. Girshick, J. Donahue, T. Darrell, J. Malik [Электронный ресурс] // Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. USA, OH, 2014. Pp. 580–587. DOI: 10.1109/CVPR.2014.81 (дата обращения: 05.01.2024).

10. Krähenbühl P., Koltun V. Geodesic object proposals [Электронный ресурс] // Computer Vision – ECCV 2014. ECCV 2014. Lecture Notes in Computer Science. Springer, Cham, 2014. Vol. 8693. DOI: 10.1007/978-3-319-10602-1_47 (дата обращения: 05.01.2024).

11. Pont-Tuset J. Multiscale combinatorial grouping for image segmentation and object proposal generation / J. Pont-Tuset, P. Arbeláez, J.T. Barron, F. Marques, J. Malik [Электронный ресурс] // IEEE Transactions on Pattern Analysis and Machine Intelligence. 2017. Vol. 39, no. 1. Pp. 128–140. DOI: 10.1109/TPAMI.2016.2537320 (дата обращения: 05.01.2024).

12. Zitnick C.L., Dollar P. Edge boxes: Locating object proposals from edges [Электронный ресурс] // Computer Vision – ECCV 2014. ECCV 2014. Lecture Notes in Computer Science. Springer, Cham, 2014. Vol. 8693. DOI: 10.1007/978-3-319-10602-1_26 (дата обращения: 05.01.2024).

13. Kuo W., Hariharan B., Malik J. Deepbox: Learning objectness with convolutional networks [Электронный ресурс] // IEEE International Conference on Computer Vision (ICCV). Santiago, Chile, 2015. Pp. 2479–2487. DOI: 10.1109/ICCV.2015.285 (дата обращения: 05.01.2024).

14. Pinheiro P.O. Learning to refine object segments / P.O. Pinheiro, T.-Y. Lin, R. Collobert, P. Dollar [Электронный ресурс] // Computer Vision – ECCV 2016. ECCV 2016. Lecture Notes in Computer Science. Springer, Cham, 2016. Vol. 9905. DOI: 10.1007/978-3-319-46448-0_5 (дата обращения: 05.01.2024).

15. Redmon J. You only look once: Unified, real-time object detection / J. Redmon, S. Divvala, R. Girshick, A. Farhadi [Электронный ресурс] // Proceedings of IEEE CVF Computer Vision and Pattern Recognition Conference (CVPR). USA, Las Vegas, NV, 2016. Pp. 779–788. DOI: 10.1109/CVPR.2016.91 (дата обращения: 05.01.2024).

16. Shanmugamani R. Deep learning for computer vision. Packt Publishing, 2018. 310 p.

17. Zhao Z.-Q. Object detection with deep learning: A review / Z.-Q. Zhao, P. Zheng, S.-T. Xu, X. Wu [Электронный ресурс] // IEEE Transactions on Neural Networks and Learning Systems. 2019. Vol. 30, iss. 11. Pp. 3212–3232. DOI: 10.1109/TNNLS.2018.2876865 (дата обращения: 05.01.2024).

18. Gatys L., Ecker A., Bethge M. A Neural algorithm of artistic style [Электронный ресурс] // Journal of Vision. 2016. Vol. 16. ID: 326. DOI: 10.1167/16.12.326 (дата обращения: 05.01.2024).

19. Changela A., Zaveri M., Verma D. A Comparative study on CORDIC algorithms and applications [Электронный ресурс] // Journal of Circuits, Systems and Computers. 2023. Vol. 32, no. 05. ID: 2330002. DOI: 10.1142/S0218126623300027 (дата обращения: 05.01.2024).

References

1. Gorbunov, A.L. (2016). Trainer of airport the specials-technician. *Civil Aviation High Technologies*, no. 225 (3), pp. 92–97. (in Russian)

2. Gorbunov, A.L. (2023). Visual coherence for augmented reality. *Advanced Engineering Research*, no. 2, pp. 180–190. DOI: 10.23947/2687-1653-2023-23-2-180-190 (in Russian)

3. Tewari, A., Fried, O., Thies, J. et al. (2020). State of the art on neural rendering. *Computer Graphics Forum*, vol. 39, issue 2, pp. 701–727. DOI: 10.1111/cgf.14022

4. Einabadi, F., Guillemaut, J., Hilton, A. (2021). Deep neural models for illumination estimation and relighting: A survey. *Computer Graphics Forum*, vol. 40, issue 6, pp. 315–331. DOI: 10.1111/cgf.14283

5. Ghasemi, Y., Jeong, H., Choi, S., Lee, J., Park, K. (2022). Deep learning-based object detection in augmented reality: A systematic review. *Computers in Industry*, vol. 139, ID: 103661. DOI: 10.1016/j.compind.2022.103661 (accessed: 05.01.2024).

6. Tu, W.-C., He, Q., Chien, S.-Y. (2016). Real-time salient object detection with a minimum spanning tree. *In: Proceedings of IEEE CVF Computer Vision and Pattern Recognition Conference (CVPR)*, pp. 2334–2342.

7. Yang, J., Yang, M.-H. (2012). Top-down visual saliency via joint CRF and dictionary learning. *In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. USA, RI, pp. 2296–2303. DOI: 10.1109/CVPR.2012.6247940 (accessed: 05.01.2024).

8. Rosin, P.L. (2009). A simple method for detecting salient regions. *Pattern Recognition*, vol. 42, issue 11, pp. 2363–2371. DOI: 10.1016/j.patcog.2009.04.021

9. Girshick, R., Donahue, J., Darrel, T., Malik, J. (2014). Rich feature hierarchies for accurate object detection and semantic segmentation. *In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, USA, OH, pp. 580–587. DOI: 10.1109/CVPR.2014.81

10. Krähenbühl, P., Koltun, V. (2014). Geodesic object proposals. *In: Computer Vision – ECCV 2014. ECCV 2014. Lecture Notes in Computer Science*. Springer, Cham, vol. 8693. DOI: 10.1007/978-3-319-10602-1_47 (accessed: 05.01.2024).

11. Pont-Tuset, J., Arbeláez, P., Barron, J.T., Marques, F., Malik, J. (2017). Multiscale combinatorial grouping for image segmentation and object proposal generation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 39, no. 1, pp. 128–140. DOI: 10.1109/TPAMI.2016.2537320 (accessed: 05.01.2024).

12. Zitnick, C.L., Dollar, P. (2014). Edge boxes: Locating object proposals from edges. *In: Computer Vision – ECCV 2014. ECCV 2014. Lecture Notes in Computer Science*. Springer, Cham, vol. 8693. DOI: 10.1007/978-3-319-10602-1_26 (accessed: 05.01.2024).

13. Kuo, W., Hariharan, B., Malik, J. (2015). Deepbox: Learning objectness with convolutional networks. *In: IEEE International Conference on Computer Vision (ICCV)*, Santiago, Chile, pp. 2479–2487. DOI: 10.1109/ICCV.2015.285 (accessed: 05.01.2024).

14. Pinheiro, P.O., Lin, T.-Y., Collobert, R., Dollar, P. (2016). Learning to refine object segment. *In: Computer Vision – ECCV 2016. ECCV 2016. Lecture Notes in Computer Science*, Springer, Cham, vol. 9905. DOI: 10.1007/978-3-319-46448-0_5 (accessed: 05.01.2024).

15. Redmon, J., Divvala, S., Girshick, R., Farhadi, A. (2016). You only look once: Unified, real-time object detection. *In: Proceedings of IEEE CVF Computer Vision and Pattern Recognition Conference (CVPR)*, USA, Las Vegas, NV, pp. 779–788. DOI: 10.1109/CVPR.2016.91 (accessed: 05.01.2024).

16. Shanmugamani, R. (2018). Deep learning for computer vision. Packt Publishing, 310 p.

17. Zhao, Z.-Q., Zheng, P., Xu, S.-T., Wu, X. (2019). Object detection with deep learning: A review. *IEEE Transactions on Neural Networks and Learning Systems*, vol. 30, issue 11, pp. 3212–3232. DOI: 10.1109/TNNLS.2018.2876865 (accessed: 05.01.2024).

18. Gatys, L., Ecker, A., Bethge, M. (2016). A Neural algorithm of artistic style. *Journal of Vision*, vol. 16, ID: 326. DOI: 10.1167/16.12.326 (accessed: 05.01.2024).

19. Changela, A., Zaveri, M., Verma, D. (2023). A Comparative study on CORDIC algorithms and applications. *Journal of Circuits, Systems and Computers*, vol. 32, no. 05, ID: 2330002. DOI: 10.1142/S0218126623300027 (accessed: 05.01.2024).

Сведения об авторах

Горбунов Андрей Леонидович, кандидат технических наук, доцент, доцент кафедры управления воздушным движением МГТУ ГА, a.gorbunov@mstuca.aero.

Ли Юньхань, аспирантка МГТУ ГА, antatanoe@gmail.com.

Information about the authors

Andrey L. Gorbunov, Candidate of Technical Sciences, Associate Professor, Associate Professor of the Air Traffic Management Chair, Moscow State Technical University of Civil Aviation, a.gorbunov@mstuca.aero.

Yunhan Li, Postgraduate Student, Moscow State Technical University of Civil Aviation, antatanoe@gmail.com.

Поступила в редакцию	22.03.2024	Received	22.03.2024
Одобрена после рецензирования	03.06.2024	Approved after reviewing	03.06.2024
Принята в печать	25.07.2024	Accepted for publication	25.07.2024