**Form Name:** Bass Connections Proposal for 2020-2021 Project Teams

|  |  |
| --- | --- |
| **Project title:** | **Deep learning for rare energy infrastructure in satellite imagery** |
| **Primary point of contact for project:** | Kyle Bradbury |
| **Which Bass Connections theme are you applying to? Please check all that apply.** | Energy & Environment |
| **Is this project a continuation of a current or previous Bass Connections project?** | Yes |

|  |  |
| --- | --- |
| **Are you also applying to connect this year-long Bass Connections project with a Data+ project in Summer 2020?** | Yes, Data+ (If you click yes, you will be asked to respond to additional questions specific to Data+) |

**2. Project Description**

|  |  |
| --- | --- |
| **Please provide brief background/context for the issue this project seeks to address. (1,500 character maximum)** | Critical information for energy access decision-making and electricity system planning is not universally available including information on village-level electricity access and reliability as well as the location and characteristics of power system infrastructure including generation, transmission, and end-use consumption. Decision makers require information to determine the optimal strategies for deploying energy resources to decide where to prioritize development and whether electrification should be accomplished through grid expansion, microgrids, or distributed generation.  Past studies demonstrated that deep learning techniques may be successfully applied to satellite imagery to fill these knowledge gaps, but these studies focused on one type of infrastructure or system characteristic and were applied in one specific geographic region, and were therefore limited in scope. Additionally, these techniques often require additional training data to adapt to new contexts or to identify rare objects and this costly process may inhibit frequent applications of these techniques at scale. |
| **Please provide a brief description of the project approach and goals, including a description of the planned research methods and activities. (4,000 character maximum)** | This project aims to develop deep learning techniques that can automatically and rapidly scan massive volumes of remotely sensed data (e.g., satellite imagery), providing detailed maps of energy infrastructure, from which critical energy systems knowledge gaps can be filled. If practically feasible, these deep learning approaches may provide a powerful new tool for researchers, policy-makers, and governments to collect energy systems information.   In past projects, we have begun to develop deep learning models that can detect energy infrastructure in satellite imagery, including power transmission lines and off-grid distributed energy generation sources such as solar photovoltaics. However, deep learning models must be trained to identify energy infrastructure using hand-labeled examples of the target objects in the real-world data. Due to the rarity of such objects, human annotators must visually inspect large quantities of imagery to identify training examples. Furthermore, this rarity tends to reduce the diversity of training examples (e.g., limited lighting conditions, background scenery), making the detection accuracy more sensitive to new conditions, which may not have been present in the training imagery.  Although the total effort to procure training imagery is several orders of magnitude smaller than manually inspecting all data (i.e., the alternative approach), this process is still relatively costly and time-consuming, limiting the value and widespread adoption of these techniques. To address this problem, we propose to generate synthetic imagery from virtual worlds, such as those in modern video games. In such virtual worlds, the precise locations of all objects are known in advance, obviating costly hand-annotation. We will create virtual worlds with various types of energy infrastructure and imaging conditions (e.g., lighting and background scenery). This will allow us to capture substantially greater quantities and diversity of training data, and do so at much lower cost, compared to conventional methods of obtaining training data.   Using these new more scalable methods of training data collection, we develop deep learning models to identify an increasing large variety of energy infrastructure to fill information gaps. Therefore, the goals of this project are threefold: (1) investigate a new form of training data procurement by using synthetically generated imagery, (2) investigate multiclass energy object identification at scale, and (3) make data for decision and policy makers in the energy access and system planning space open source and available for two cities. |
| **What are anticipated outcomes from this project? (e.g., publications, website, app, data collection for further research/grant) (1,000 characters maximum)** | Through the course of this project (the Bass Connections project and connected Data+ summer experience), the team will:  1. Develop a new multiclass dataset of rare energy infrastructure objects and other rare object types  2. Create a tool for the synthetic generation of overhead imagery containing rare objects for improving the automatic detection capabilities of those objects; this tool will be shared on github after publication on the idea  3. A team poster and final presentation open to the public  4. A conference or journal paper based on the methods and/or datasets developed |

**3. Team Composition**

|  |  |
| --- | --- |
| **Team Leaders: Please list all team leaders below, including Name, Title, and Department/School.** | Kyle Bradbury Managing Director, Energy Data Analytics Lab and Lecturing Fellow Energy Initiative and Pratt, Electrical and Computer Engineering  Jordan Malof Assistant Research Professor Pratt, Electrical and Computer Engineering |
| **Team Contributors: Please list all team contributors below, including Name, Title, and Department/School.** | Leslie Collins Professor of Electrical and Computer Engineering, Professor of Biomedical Engineering Pratt, Electrical and Computer Engineering  Robert Fetter Senior Policy Associate, Nicholas Institute; Postdoctoral Research Fellow, Energy Initiative Energy Initiative and Nicholas Institute  Marc Jeuland Associate Professor, Public Policy and Global Health Sanford School of Public Policy and Duke Global Health Institute  Luana Marangon Lima Visiting Assistant Professor Nicholas School of the Environment, Environmental Science & Policy Division  Robyn Meeks Assistant Professor, Environmental and Energy Economics Sanford School of Public Policy |
| **PROJECT MANAGER: Do you plan to assign someone other than a faculty leader as a “project manager” for your team?** | Yes |
| **STUDENT PARTICIPATION: Ideally, how many graduate students would you select to participate on this team? (Numeric responses only, please)** | 2 |
| **Ideally, how many undergraduate students would you select to participate on this team? (Numeric responses only, please)** | 6 |
| **What would be the ideal composition of team members for this project? What majors, disciplines, skills, backgrounds, or perspectives would you like to have on the team? (1,200 characters maximum)** | The ideal composition of our team would be six to nine students mixed between undergraduates and graduates from disciplines including engineering, computer science, economics, public policy, environmental science, mathematics, and statistics. Students with some experience in quantitative methods are preferable. Experience with programming is required, and experience with Python programming is beneficial. Graduate students in this project will take on a leadership role as project managers. Ideally, we would like to engage one to two graduate students that collectively have excellent management skills and preferably strong technical skills or a willingness to get up-to-speed rapidly with respect to technical skills. |
| **EXTERNAL PARTNERS: Will your team also include any external organizations or individuals as either partners, clients, study subjects, beneficiaries of the work, etc.?** | Not sure yet |

**4. Project Approach**

|  |  |
| --- | --- |
| **Team approach: How will you facilitate collaborative research on the team? How often and in what format will the team meet? How will you divide tasks? How you will ensure effective management of the project (e.g., appoint a student as a project manager, assign that role to a faculty leader, etc.)? (1,500 character maximum)** | Most students enter a Bass Connections project team with minimal research or team experience, and through our project team we seek to provide these students the opportunity to grow in both. We employ an intensive two-semester schedule that includes twice-weekly meetings with team leaders. We begin with team leader-guided content dissemination and gradually transition to student-facilitated team meetings with weekly status updates. There are numerous team building activities including an escape room activity as well as a StrengthsFinder assessment to understand relative strengths of each team member. Additionally, the first semester contains a project sprint to help the team further cohere. Participating graduate students will serve as project managers for the team. |
| **Student opportunities: What might students gain from their participation (e.g., conducting research directly with subjects, contributing to publications, using language skills)? What unique and differentiated learning opportunities would be available for graduate students? (1,500 character maximum)** | By the end of this team project experience, you should be able to...  - Recognize improvements in your verbal and written communication skills for technical and nontechnical information  - Create cogent presentations that succinctly describe your work and its key outcomes  - Engage in team-based problem solving and empower you to take ownership of project management and organization  - Describe the relationship between access to electricity, economic well-being, and health  - Apply a formalized research process including problem definition, literature review, research project design, execution, analysis, and interpretation  - Apply advanced data analytics tools to energy data, including computational tools such as the Python programming language and deep learning algorithms  - Explain the fundamentals of machine learning, pattern recognition, and computer vision  The graduate students will be project managers. The project manager will attend all team meetings and is responsible for the day-to-day work of the project. She will help to guide and support team member needs, identify topics that the team could use help with, and help to keep the team on track with deadlines. She will also help to guide the project's research direction. |
| **Timeline: You may upload a file here.** | <https://www.dropbox.com/s/w5df60drqk2mq8n/Deeplearningforrareenergyinfrastructureinsatelliteimagery_81765283_EnergyDataAnalyticsLabBassConnectionsProposal20191104.pdf?dl=0> |

**5. Project Details**

|  |  |
| --- | --- |
| **Does your proposal include travel for students beyond the Triangle?** | No |
| **Does your project proposal include opportunities for students to conduct research during the summer?** | Yes, Summer 2020 |
| **Please include some additional information about the nature of the summer research opportunity:** | Students may (but are not required to participate in) Data+ in summer 2020 (details about Data+ in the related questions). Students would need to apply separately for Data+ and Bass Connections since the Data+ project would not be able to accept all of the Bass Connections student team members during the summer. |
| **Will your team be interacting with minors through this project?** | No, we do not plan to interact with non-Duke minors |
| **Does this project relate to any in-progress invention disclosures with the Office of Licensing and Ventures (OLV )?** | No |

**6. Budget Estimate**

|  |  |
| --- | --- |
| **Total Budget Request (numeric response only, please):** | 18,460 |
| **You may upload a budget table here, or you may complete the fields below.** | <https://www.dropbox.com/s/3gta5z8211n3w05/Deeplearningforrareenergyinfrastructureinsatelliteimagery_81765287_Budget.pdf?dl=0> |

**Additional Questions for Continuing Projects**

|  |  |
| --- | --- |
| **You noted in section 1 that this proposal relates to a current and/or previous Bass Connections project. Please provide the name of the project to which this proposal relates.** | A Wider Lens on Energy: Adapting Deep Learning Techniques to Inform Energy Access Decisions (2019-2020) |
| **How do you anticipate that this proposed project will build on your current and/or previous related Bass Connections project? How will it differ? (1,200 character maximum)** | This year (2019-2020) we're exploring how synthetic imagery could help increase the quantity and diversity of training imagery for deep learning models for geographic domain adaptation. We're proposing to expand this concept to be able to identify rare objects (as most energy infrastructure are) so that we can both detect energy infrastructure in new locations and in places with limited training examples. As we continue to expand the capabilities of our recognition algorithms, the collection of useful training data has become an important bottleneck, prompting us to explore the use of synthetic overhead imagery. Supported by this year's Bass Connections and Data+ teams, we have developed seminal approaches for generating useful synthetic overhead imagery for training deep learning models. In this ongoing work we are using computer generated models of energy-related infrastructure objects to populate our virtual worlds. We hypothesize this research will increase the scalability and applicability of these automated energy infrastructure assessment techniques, which are critical for addressing data gaps in identifying pathways for electrification in low-income countries. |
| **What are the most important outcomes to-date of your current and/or previous related Bass Connections project? (1,200 character maximum)** | Our current project has built a new synthetic imagery dataset for helping to adapt to new geographic domains (especially those where energy access is increasing) through Data+ that will be expanded through Bass Connections this year and will work towards a paper related to geographic domain adaptation. Past projects used remote sensing and satellite imagery data to estimate the generation capacity of solar photovoltaic arrays (2015-2016), energy consumption of buildings (2016-2017), and energy access from remote sensing data (2017-2018), and identify transmission lines in overhead imagery (2018-2019). These projects led to 12 conference and journal papers and 5 (soon to be 6) publically-available datasets for the research community. Additionally, students on these teams have won multiple poster competitions (internal and at regional conferences). Most importantly, the team leaders of this project have received notes from students every year of how participation in this team was a key factor in receiving a job or internship opportunity. |
| **What challenges has your current and/or previous team encountered and how would you address any such issues if funded next year? (1,000 character maximum)** | The perennial challenge faced in Bass Connections teams is how to keep the students level of motivation and engagement constant throughout the two semesters. It has the tendency to ebb and flow with varying responsibilities. In past teams, this had led to periods of short bursts of intense activity before deadlines. We will address this issue by requiring more frequent report outs for student progress, setting clearer and measurable objectives for the team each week (with the team participating in that goal-setting exercise to learn their capabilities and limits), and to make this a graded component of the experience so students have a strong incentive to put in the time and effort. |

**Data+ Questions (Only applies to teams applying jointly to Data+)**

|  |  |
| --- | --- |
| **What would the goal of the Data+ project be and how would you envision the Data+ team’s work connecting with the goals of the year-long Bass Connections project?** | Over the course of this summer experience, students on this Data+ project team will lay the foundation for the Bass Connections team by creating a labeled, synthetic energy infrastructure remote sensing dataset containing rare objects. The energy infrastructure we will synthesize will be (1) generation, (2) transmission and (3) rare end uses (such as electric vehicles). This dataset will be comprised of synthetic representations of these objects to identifying rare objects in imagery (such as substations and most transmission infrastructure). The Bass Connections team will use these data as the foundation of their research. |
| **What tangible product(s) would you envision the Data+ team creating? (e.g., well-commented software, visualization device, or a detailed curation of previously raw data).** | This Data+ project will curate a new machine learning dataset that contains synthetic imagery for key, rare energy infrastructure including (1) generation, (2) transmission, and (3) rare end-uses. This dataset will be fully designed from synthetic imagery, by comparing it to real imagery and validating that the performance of object detection is enhanced by the particular synthetic models. We will also release code for generating these synthetic data which will enable the broader use of this research. |
| **Please describe the nature of the dataset(s) that will be analyzed by the student team including: what’s in it, how was it collected and for which purpose, how large it is, etc.** | For the reference real-world data, we will use a diverse collection of publically available imagery for data curation with this project including USGS high resolution orthoimagery data, a vast database of publicly available satellite and aerial photography data from the USGS website, as well as data from Africa, Southeast Asia, Europe and the Caribbean. The USGS data range in vintage from 2002 through 2014, but have a resolution as high as 0.3 meters for some cities, and the data from other parts of the world are provided primarily through the Maxar Open Data program at 0.3 or 0.5 meter resolution with some of the data as recent as 2018. These data are large in size, with one city representing at least 20 to 30 Gb of data. Students will be required to download and manipulate the data, to expose them to best practices for managing and processing large data sets. |
| **For each dataset, indicate whether it is owned and/or is being provided by an outside party. If so, please describe the intended path towards ensuring that students will be granted the ability to access the dataset (we are often able to assist in crafting Data Use Agreements with outside parties, for example).** | No IRB approval required. |
| **For each dataset, indicate whether you anticipate IRB approval will be needed for student access, and if not, why not. If IRB approval will be needed, indicate whether a protocol already exists, and your plan for incorporating the student involvement. If it does not already exist, please describe your plan (including a timeline) for obtaining one.** | All datasets are publicly available. |
| **Do you already have a project manager in mind for this Data+ project? (Typically a PhD student who mentors the undergraduate students about 4-5 hours a week)** | We wish to recruit a Ph.D. student or advanced masters student for this position and if both Bass Connections and Data+ projects are approved, we would like the graduate student to work full time throughout the summer with the team, with support provided from the Bass Connections budget request (if approved). |
| **Who would be the primary point of contact for the Data+ project, if funded? (This could be the same as for the year-long Bass Connections project, or different)** | Kyle Bradbury. Managing Director, Energy Data Analytics Lab |